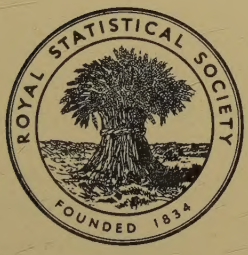
APPLIED STATISTICS

A Journal of the
Royal Statistical Society

EDITED BY
LEONARD H. C. TIPPETT



VOLUME TWO

Oliver and Boyd Ltd.
London: 98 Great Russell Street, W.C.1
Edinburgh: Tweeddale Court

1953

CONTENTS

No. 1. MARCH

Page

The Census of Distribution .. .. .. *Dudley A. Clark* 1

Estimating Variability from the Differences between Successive Readings (with Appendix: Some Theoretical Aspects *Herman O. Hartley*) .. .. *Joan Keen and Denys J. Page* 13

Short-term Fluctuations in Petroleum Consumption *John Douglas Butler* 24

A Statistical Approach to a Problem of Chemical Manufacture *Andrew Muir and Frank Downton* 30

A Comparison of Serial Number Digit Sampling with Systematic and Random Sampling .. *Orville O. Price and Alfred J. McKeon* 39

Stochastic Processes or the Statistics of Change *Maurice S. Bartlett* 44

Questions and Answers 65

Notes and Comments 67

Industrial Applications Section of the Royal Statistical Society: List of Meetings 68

Book Reviews:

The Design and Analysis of Experiments, Oscar Kempthorne 70

Statistics and their Application to Commerce, A. Lester Boddington (and A. R. Ilersic) 70

The Sources and Nature of the Statistics of the United Kingdom Volume I, M. G. Kendall (Editor) 71

The Methods of Statistics, L. H. C. Tippett 72

No. 2. JUNE

Experiences of Correlation Analysis .. .. *Donald G. Beech* 73

Economic Forecasting in Great Britain .. .. *Eric Shankleman* 86

Statistics and Prediction .. .. .. *Cyril Davenport Hughes* 101

The Origin and Development of Factor Analysis *Douglas F. Vincent* 107

The Economics of Sequential Sampling Procedures for Defectives *David G. Champernowne* 118

Letter to the Editor: Estimating Individual Expenditure from Family Totals 131

Notes and Comments 133

Industrial Applications Section of the Royal Statistical Society: List and Reports of Meetings 136

Study Section of the Royal Statistical Society: Reports of Meetings .. 138

Study Circle on Medical Statistics: Royal Statistical Society 140

to the need for more complete information and therefore strengthened the case for a distribution census. Indeed, the committee appointed in 1945 by the Board of Trade to enquire into the case for a census considered this aspect of such importance that in its report[3] it recommended that 'in the general public interest and in particular owing to the statistical needs connected with employment policy, a Census of Distribution should be taken in 1949 and at least once every five years thereafter.' The committee made recommendations about the field to be covered by the Census and suggested that a pilot enquiry should first be conducted by the Board of Trade in order to test the most suitable methods for the full census.

The Pilot Enquiry was originally planned to take place in 1947. It was clear, however, that certain sections of the distributive trades so strongly opposed the idea of a census that a voluntary enquiry, even if limited to a few test areas, would encounter a great deal of criticism. The opposition to the Census was based on the view that it was an unwarrantable and unnecessary intrusion into the private affairs of individuals or firms: in the background was the conviction that the results would be twisted to support a preconceived case for further government regulation. The strength of the opposition and the fuel crisis of 1947 combined to cause a postponement of the Pilot Enquiry until 1948 and the full Census until 1950. No results of the Pilot Enquiry have been published, but it is believed that the response was very poor.

Meanwhile, statutory authority for the full census had been established by the Statistics of Trade Act, 1947. In July 1948 the President of the Board of Trade set up a consultative committee to advise on the arrangements for the full Census. In January 1949 the Census was postponed for a further year on the grounds that there was not enough time to make all the necessary arrangements: it may well have been that political considerations also affected the decision for a second postponement. At the beginning of 1951, however, the Census went ahead in spite of some continuing opposition.

This brief review of the history of the Census illustrates very clearly the difficulties of carrying out an operation of this kind when there is strong opposition. It is true that by no means all traders have opposed the Census: some distributors' organisations have given considerable assistance to the Census Office in preparing questionnaires and in explaining the Census to their members. But the existence of a vocal opposition has affected the value of the Census in a number of ways. It will undoubtedly affect the quality of the returns and has led to delay and additional expense in collecting them. And, perhaps most important of all, it was bound to affect the decision as to how much information could be asked of traders: in cases of doubt there was an inevitable tendency to omit a particular question rather than run the risk of antagonising the opposition still further.

Some opposition to a first census was, of course, inevitable, especially

in a nation where form-filling of any kind is unpopular. It does appear, however, that some special measures are needed to 'put over' the idea of a census in terms of the practical value of the results to the business community.

We turn now to a brief description of the Census under three headings: (1) the field covered, (2) information obtained, and (3) the planning and operation of the Census and safeguards against disclosure of individual returns.

The Field Covered by the Census

The problem of defining the field to be covered in the Census illustrates very clearly the enormous complexity of the processes and operations which are loosely regarded as 'distribution.' One of the most serious difficulties lies in deciding how many of the service trades, such as transport, hairdressing, shoe-repairing, estate agencies, hotels, and catering establishments, should be included under the heading of distribution. Another problem is setting the dividing line between manufacturing and distribution: should the selling functions (e.g. advertising) performed by a firm primarily engaged in manufacturing be covered under the Census of Distribution or the Census of Production ?

As a general rule, the field to be covered by the Census of Distribution was defined as all establishments primarily engaged in selling goods, either to the public or for resale, and those engaged in certain specified service trades. The latter included non-residential catering establishments such as cafés, snack bars, and canteens, but not public houses, clubs, hotels, or catering in hospitals or schools. Hairdressers, portrait photographers, funeral directors, pawnbrokers, shoe and other repair businesses, and warehousing for the trade, were included. The decision to exclude hotels, public houses, clubs, etc., was taken on administrative grounds in order to keep within the capacity of the Census Office: it is possible that they will be covered in a future census or by a special enquiry before the next general census. Apart from certain minor exceptions, establishments primarily engaged in manufacturing were excluded, although if they ran a canteen a return had to be made in respect of the latter.

A fair idea of the field covered can be obtained from the list of different types of questionnaire which were designed for particular classes of trader. These were as follows:

Wholesalers (including import merchants, produce brokers, selling agents, warehousing firms, etc.).
Large retailers (with trading of £5000 a year or over).
Small retailers.
Multiple retailers ('multiple' being defined as a business with ten or more branches).
Co-operative societies.
Stallholders, street traders, and pedlars.

Large motor traders (motor dealers, filling stations, repair garages, etc., with trading of more than £5000 a year).
Small motor traders.
Large caterers.
Small caterers.
Multiple caterers.
Canteens.
Pawnbrokers.
Service traders (undertakers, portrait photographers, etc.).
Boot and shoe repairers.
Other repair businesses.

Information Obtained

The information required from trades varied somewhat between the different categories listed above. The forms for small traders were much simpler than those for larger businesses. There were, however, certain fairly well defined types of information collected in varying amounts of detail from every business. These may be grouped under the following headings:

Kind of Business. The trader was asked to give his own description of the business: a retailer, for example, might classify himself as a grocer, butcher, ironmonger, or baker. 'Wholesalers' were classified into a large number of sub-groups, including wholesalers in the normal sense, merchants of various sorts, agents, brokers, warehousing firms, etc.

Types of Goods Handled. Although the trader's own description of his business gives a guide to the type of commodity which he sells, it is only a very rough indication. Some grocers, for example, sell tobacco goods; others do not. 'General stores' may vary greatly in the kind of goods handled. In order to make possible a more accurate and objective classification of businesses, information was collected on the categories of goods actually sold, the amount of detail collected under this heading varying between one type of distributor and another. The small retailer, for example, was required to indicate only the four main groups of goods which he handled, in order of importance; these main groups were fairly broad ones, such as 'groceries and provisions,' 'chocolate and sugar confectionery and ice-cream,' and 'men's and boys' shirts, underwear, hats, ties, etc.' The wholesaler, on the other hand, was asked for more detailed information about each commodity group and was also required to say whether, in each group, he sold a 'general range' or only certain specific articles.

Sales. The amount of information collected under this heading also varied considerably between one type of trader and another. Small businesses were required to state total takings during the census year (either the calendar year 1950 or the business year ending between 5th April 1950 and 6th April 1951) and to estimate a division of takings between sales of goods (including those taken for the proprietor's own use), sales of meals and refreshments, and charges for

services. They were not expected to analyse sales by commodities. Large retailers were, however, required to provide not only a breakdown of sales into receipts from sales of goods, meals, and services respectively, but also an analysis by broad commodity groups. The instructions covering this analysis stated that 'reasonable estimates' would be acceptable if no detailed records were available. Multiple retailers, co-operative societies, and wholesalers, were required to analyse their sales according to the classes of customer to which they were made (to retailers, direct to the public, to other wholesalers, to buyers abroad, etc.). Wholesalers were also required to indicate the proportions of total sales represented by each of the four main groups of commodity sold.

Purchases. With the exception of co-operative societies, where an analysis by departments was required, no attempt was made in any of the questionnaires to obtain data on the value of purchases of different types of commodity. It follows from this that the census results will give very little information on gross margins received by distributors on different classes of goods. Most traders were asked simply for the total amount spent on purchases of goods for resale and materials used for repairs, etc.; purchase tax was included in this figure. Slightly more detail was required from wholesalers, including separate figures for the value of goods sold on commission. Multiple firms were asked for purchase data for the firm only and not for each branch separately.

Stocks. As for purchases, no analysis of stocks by commodities was obtained in the Census. Total value figures for the beginning and end of the trading period were required, the basis of valuation being cost or market value, whichever was the lower.

Employment. A detailed analysis of people working in the business was obtained from all types of distributors for the week ending 24th June 1950. Working proprietors, members of the proprietor's family or friends, and employees paid a regular wage, were recorded separately and each of these groups was analysed by sex and by full or part time. For employees there was an additional breakdown into adults and juveniles.

Expenses. The only expense items on which information was collected from all classes of distributors were wages and salaries for the specified week to which the employment figures related and also for the census year. Larger businesses were required to give separate figures, on a yearly basis, for bonuses, commission, etc. Wholesalers were asked for details of their transport costs, broken down into expenses of operating their own vehicles and payments to contractors for carriage outwards.

Other Information. In addition to the main categories of information listed above, the questionnaire used for some types of distributor contained certain questions on matters which were not generally applicable to all traders. For example, the co-operative societies were asked for details of dividend payments; some large businesses were required to

state the amounts receivable from the Ministry of Food and other government bodies in respect of subsidies on bread-baking, milk handling, etc. Multiple firms were asked for separate details of the numbers employed in their head offices. The forms for caterers and canteens also contained a number of questions applicable only to that type of business, such as payments in kind made to employees.

From the nature of the data listed above it is clear that the census results will produce a fairly complete picture of the number of distributors of various types, of the employment situation in the distributive trades, of the total values of sales, purchases, and stocks (and therefore of gross profits), and of wages paid. The Census will give very little detailed information on margins obtained on specific commodities, on the number of intermediaries through which different types of goods pass on their way from manufacturer to consumer, on the number of outlets for particular kinds of widely distributed goods, or on expenses (other than wages) incurred by distributors.

The Planning and Operation of the Census

The Statistics Division of the Board of Trade is responsible for running the Census. The administrative work is handled by a Census of Distribution Office within this Division.

The first stage in the census field work was an enumeration of the addresses to which forms should be sent and of the type of form appropriate for each address, where this could be ascertained in advance. Several methods of making this enumeration were tried in the Pilot Survey, but most of the work was eventually done by members of the staff of the Census Office. The Post Office, other government departments, and local authorities were probably able to assist in providing lists against which the enumeration could be checked.

A preliminary analysis of the results of the enumeration has been published in a booklet called 'Britain's Shops.'[4] This gives a geographical analysis, by counties and large towns, of retail shops and service trade establishments classified according to the enumerator's assessment, judged by the appearance of the premises, of the business being conducted. It shows a total of 551 000 shops and 163 000 service establishments in Great Britain, the largest single group being 'grocers and dairy shops' (134 000). In the Census Report the classification of shops will be based on examination of the particulars supplied by traders themselves and not on the enumerator's opinion. In spite of the inevitable inaccuracies in enumerators' classifications, however, the publication of this preliminary analysis has served a useful purpose.

The census forms were sent out by post early in 1951. When they were received back they were passed through the usual processes of checking, coding, and preparation of the punched cards from which tabulations are being made. The Board of Trade announced that they hoped to have substantial results available before the end of 1952, but in fact only the preliminary booklet had been published.

Certain provisions were made in the Statistics of Trade Act, 1947, with regard to the security of returns made by traders. Disclosure of individual returns without the previous written consent of the person carrying on the business was prohibited except (*a*) for the purpose of proceedings for an offence in connection with the Census, and (*b*) when authorised by the President of the Board of Trade for the use of another government department or the Import Duties Advisory Committee in connection with the functions of that department or committee. A later Order made in 1950 annulled the second exception so that the only grounds for disclosure of individual returns are now proceedings in connection with the Census itself and returns made under the Census.

The question of possible disclosure of information about an individual business will, however, have a bearing on the tabulation of the final results. Very detailed analyses of, say, types of outlets in particular towns might reveal, directly or by inference, information about a particular business. This point will doubtless be carefully watched.

The Value of the Census Results

Data on the structure and operation of the distributive trades are relevant to a very broad range of problems. Producers need information about the channels through which they may market their goods. Trade unions are interested in the employment situation in distribution. The economic statistician requires data on the contribution of the distributive sector to the national output. Town planners need facts on the existing pattern of distribution in particular localities. The government and all those concerned with broad issues of social and economic policy need more details than are available at present about the place of the distributive trades in the national economy.

Until the results of the Census have been published and assimilated it will naturally be impossible to make any final assessment of their value. The only really convincing proofs of the usefulness of the results will be case-studies showing how they have been applied to practical problems. At this stage the problem of whether the results will justify the time and money spent in getting them can be discussed only in general terms.

It must be emphasised that the aim of a census is to provide background information: it is analogous to basic rather than applied research in the physical sciences. In the nature of things it is absurd to expect that the Census will provide ready-made answers to problems such as whether it would be in the national interest to reduce the number of shops dealing in particular commodities. It is particularly unfortunate that, in popular discussions of the Census, some of its supporters have tended to give the impression that when the results are available it will then be possible to judge whether 'distribution costs too much' (whatever that phrase may mean). The value of the Census must consist primarily in the fact that it will provide a framework

from which more detailed studies of practical problems may be developed on scientific lines.

The problems which the Census should, in this sense, help to solve may be grouped into (1) broad questions of social and economic policy and (2) practical marketing problems encountered by businessmen.

In the sphere of broad economic and social problems the census results will have two kinds of application. Firstly, the data should bring some factual evidence to bear on the much-discussed problem of the efficiency of present distributive arrangements. Secondly, the results will contribute valuable material for the estimates of national income and output and other economic statistics which now form the basis of many major decisions on economic policy. In neither case will the Census provide all the information that is necessary; but in both cases it will give a starting point from which further research will be possible.

The data which will be available on distributors' profits, which have applications to both kinds of problem, illustrate this. In general, the only information produced by the Census will be on gross profits made in different types of business during the census year. There will be no information on net profits or on the margins obtained on specific commodities except where a particular commodity is sold almost exclusively by specialist outlets handling nothing else; gross profits on total trading will then be almost the same as gross profits from the commodity in question. In general, however, there will be little if any information on the gross and net margins obtained by different categories of distributors on particular kinds of goods. The overall total figures of gross profits will nevertheless be of value. They will provide a much sounder basis for discussion on the cost of distribution than the isolated examples of particular commodities which are too often used to support general conclusions on one side or the other. The total figures will also be useful as a check on estimates of the value of the 'output' of distributive trades which are incorporated in the national income statistics. These statistics now form the basis of important social and economic decisions, but at present there are no means by which the outsider can judge the margins of error to which they are subject. This is a serious weakness and detracts considerably from their value. The magnitude of revisions which have been made from time to time suggests that some of the margins of error are substantial. The use of the census results as a broad check on the part of the estimates which relate to the distributive sector will therefore serve an important function.

The greatest value of the Census will, however, be in the facilities which it will provide for more detailed sample studies of particular aspects of the problems of distribution. Cost studies based on the census framework would be of great value; so also would studies aimed at producing a basis for more reliable measures of trends in retail trade,

the level of distributors' stocks, etc., than are at present available. Information on points such as these would be valuable both in connection with national income statistics and in the preparation of other short-term measures of trends in the economic situation.

Business Uses of Census Data

The businessman is materially affected by decisions made on the broad plane of social and economic policy. Indirectly, therefore, he is interested in facts that can be brought to bear on such problems of policy and in the correct interpretation of those facts. As regards the more direct uses of the census results in running a business, the value obtained will vary considerably from one type of business to another. A firm selling heavy engineering equipment, where each order is probably negotiated directly with the customer, will find little of value. On the other hand, the manufacturer of consumer goods selling in a wide market may find a great deal of useful information. Large-scale distributive organisations should also find valuable data; the uses of the results to small traders is more problematical, but the census results should be of considerable interest to the trade associations that are concerned with the problem of the small distributor's place in the national economy.

It is impossible in this paper to discuss in detail the value of the census results to all types of business. It is therefore proposed to limit the discussion to some examples of the ways in which the census data may be of value to manufacturers in marketing their goods and to the distributive trades themselves.

With a few exceptions, manufacturers do not sell their goods direct to the public: they rely on a chain of distributive arrangements over which, in most cases, they have little control. They must therefore adapt their products and their selling methods to existing distributive arrangements. Failure to achieve this adaptation can be very costly and may lead to serious waste of resources. To take an extreme example, it would be wasteful to spend money advertising a product of the kind that is usually bought in small general stores if the saleman's efforts had been concentrated on selling the product to a relatively small number of specialist dealers. It is reasonable to suppose that a manufacturer with a fair knowledge of the trade would avoid obvious mistakes such as this. But the problem of adaptation to the existing distributive arrangements may involve more subtle difficulties in the solution of which a more detailed knowledge will be required.

A suitable design of the product and its packaging will, for example, be dependent on the kind of distribution channels through which it reaches the consumer. If a substantial proportion of the sales of a particular commodity is through 'self-service' stores, special attention may have to be given to the effectiveness of the package as an aid to sales. The extent to which the manufacturer relies on intermediaries to distribute his goods will depend very largely on whether the

commodity is sold mainly through specialist outlets, each with a fairly large volume of turnover, or through a large number of small outlets. There will be a level of turnover at which direct supply by the manufacturer to the retailer is economic; for outlets where turnover does not reach this level it will probably be more economical to arrange supplies through a wholesaler.

Once he has decided on the level of turnover at which a direct account becomes economic the manufacturer wants to know where, and in what kind of outlets, he should try to get new business. It then becomes necessary to compare the present density of direct accounts with some reasonable standard; for this purpose the manufacturer needs to know the total number of outlets handling the type of goods in which he is interested, their area distribution, and an analysis by types of outlets and by turnover-groups. If these figures are available, a comparison with the pattern of existing direct accounts may show, for example, that whereas in one area the manufacturer has direct accounts with almost all the retail outlets with a turnover exceeding that regarded as a minimum for direct supply, in another area the proportion is less than half. This state of affairs may reveal certain weaknesses in the location and training of salesmen and a need for special efforts to promote sales in the area concerned.

The completeness of the picture which the census results will give of the structure of the distributive trades will vary greatly from one group of commodities to another. For goods such as boots and shoes, which are handled mainly by specialist outlets or large distributors, the Census should provide reasonably accurate data on such points as the total number of outlets, their geographical location, the number of wholesalers and retailers, the value of sales, stocks and their duration, and gross profits earned. For goods handled mainly by general shops the data relating to a particular commodity will be very much less complete. It will be impossible to estimate the total number of outlets, mainly because of (1) the lack of detail about commodities handled by small traders, (2) the broad nature of the commodity groups under which information on types of goods handled was collected, and (3) the fact that the commodity may be sold in outlets such as public houses, which were excluded from the census coverage. For the same reasons it will also be impossible to obtain data on total sales, stocks, and margins, for a particular commodity.

These limitations do not mean, however, that there will be nothing of value in the census results to those interested in this type of commodity. In some cases the census data will cover a large proportion of the distributive channels even if it does not cover them completely. Many of these limitations were inevitable in a first census; indeed, other countries, with considerably more experience than our own in this field, strongly support the view that a first census should be kept as simple as possible. It is important, however, that the gaps resulting from policy decisions as to how much it was reasonable to attempt in

this census should be realised, in the hope that improvements may be made when future work is being considered.

The possibility of making these improvements is bound to depend on the attitude of the distributive trades towards future censuses. It is therefore most important that all sections of the trade should be convinced of the value of the results of the first census, and there are a number of very important points of which the present Census will provide data of special interest to the distributive trades themselves. In the first place, the information gained on gross profits should give some factual basis for discussions of the cost of distribution in relation to the services performed. There is, at present, a great deal of discussion, but there are very few facts on this point. Consequently the subject has become bedevilled with sweeping and quite often meaningless generalisations. It will surely be in the long-term interests of the distributive trades to encourage a more factual approach to the problem. The census results will provide only a few basic facts—details will need to be filled in later. But they will at least provide a basis for a scientific study.

Secondly, the census results will provide each distributor with some kind of yardstick against which he can measure his own performance in such matters as gross profits per unit of turnover, unit wage-costs, stock durations, etc. This will also be the case for groups of businesses, such as departmental stores or co-operative societies, who will be able to compare their own position with those of businesses organised in different ways.

Thirdly, distributive undertakings looking for ways to expand their business will be able to use the census results in much the same way as manufacturers. Wholesalers, for example, will find it valuable to know the number of outlets in their localities for the commodities in which they deal.

Conclusions

The object of this paper has been to suggest certain immediate uses of the census results, both in broad economic policy-making and in business decisions. The limitations of the data imposed by the scope of the Census and of the information obtained have been emphasised. When plans for the next Census are being considered, every attempt should be made to fill in the more serious gaps. But we must not lose sight of the fact that the Census is primarily a piece of fundamental research which should serve as a basis for more detailed and intensive studies of particular parts of the field. Many of the gaps in knowledge of the structure and operation of the distributive trades could be filled by sample enquiries based on the census framework. The Census of Distribution Committee recommended in its report[3] that 'to enable short-term trends to be followed in intervening years, the results should be supplemented by sampling.' The purpose of sample enquiries should be not only to measure trends over a period of time but also to examine

specific problems, such as the analysis of distribution costs which it would be difficult, if not impossible, to cover in a full census.

Many commercial organisations, as well as bodies engaged in social and economic research, may wish to make more intensive studies of special aspects of distribution. Within the limits imposed by the confidential nature of individual returns it is to be hoped that additional tabulations or data required for sampling purposes will be made available.

REFERENCES

[1] *Distribution Censuses. An International Study.* International Chamber of Commerce Brochure No. 154.

[2] *Trial Census of Distribution in Six Towns.* British National Committee, International Chamber of Commerce.

[3] *Report of the Census of Distribution Committee—March* 1946. (Cmd. 6764), HMSO, London.

[4] *Britain's Shops.* 1952, HMSO, London.

ESTIMATING VARIABILITY FROM THE DIFFERENCES BETWEEN SUCCESSIVE READINGS

JOAN KEEN and DENYS J. PAGE*

Communication from the Staff of the Research Laboratories of The General Electric Co. Ltd., Wembley, England

The method described in this article is not only convenient for use in control chart work but it also allows control chart procedures to be extended easily to situations, such as often arise in practice, in which the ordinary procedures do not apply. Miss Keen and Mr Page give several practical examples in the article and there is an appendix by Dr Hartley dealing with some theoretical points.

A RAPID estimate of the standard deviation of a distribution that is nearly normal can be obtained by regarding the mean successive difference between adjacent readings, arranged in order of collection, as an approximation to the mean range in samples of two. This method was suggested by Mr W. J. Jennett for control chart work carried out in the MO Valve Company Ltd. in 1942, and has since been used extensively in The General Electric Company Ltd. It is only recently, however, that the authors were encouraged by Dr H. O. Hartley to collect together several practical examples of the uses of the method under slightly differing circumstances, and he comments on these uses in an appendix to this article.

The calculation of the standard deviation is made in three steps:

(i) take the differences between the consecutive values,
(ii) average the differences (ignoring sign), and
(iii) regarding them as ranges in random samples of two, divide the mean difference by the factor $d_2 = 1{\cdot}128$.†

For example, for the efficiencies of the 20 tungsten lamps listed in Table I, the standard deviation estimated by the normal root-mean-square method is 0·38. The calculation by mean successive difference gives 0·40, which for all practical purposes is in good agreement.

Single Estimates

Common types of record chart are those on which, for example, percentages of scrap material or measurements of a slow-moving, or large-scale chemical, process are plotted. In these cases the points plotted are single determinations, not sample averages, and it is not

* Now with the Southern Electricity Board.

† This is taken from Table 3 of *Quality Control Chart Technique when Manufacturing to a Specification*, by B. P. Dudding and W. J. Jennett. The Research Laboratories, The General Electric Co. Ltd., England, 1944.

TABLE I

Efficiencies of electric lamps (lumens per watt) and estimation of standard deviation from successive differences

Efficiency x	Successive Differences	
10·3		
	0·8	
9·5		
	0·2	
9·7		
	0·3	
10·0		
	0·3	
10·3		Mean successive difference
	0·2	$\bar{w}_x = 8{\cdot}5/19 = 0{\cdot}45$
10·1		
	0·4	
10·5		
	0·7	Estimated standard deviation
9·8		
	0·8	$= 0{\cdot}45/1{\cdot}128 = 0{\cdot}40$
9·0		
	0·7	
9·7		
	0·3	
10·0		
	0	
10·0		
	0·4	
9·6		
	0·5	
10·1		
	0·1	
10·0		
	0·3	
9·7		
	0·6	
10·3		
	0·2	
10·1		
	0·7	
9·4		
	1·0	
10·4		
	8·5	

always evident how to prepare a control chart. The nearest equivalent to the usual within-sample range chart is a plot of the difference between successive determinations, as for the percentages of second grade crockery in Fig. 1, and this may be used to estimate the short-term variability against which changes with time can be judged. For the data of Fig. 1 the standard deviation, using the method already outlined, is 4·4 per cent., and the fluctuations outside the 'two-sigma' limits show that there are some changes in level with time.

The percentage of second-grade crockery is to a large extent controlled by the operator concerned with any given process in its manufacture, and he is much more likely to be able either to maintain or improve his level of quality if he can see how much bad work he is normally producing. Fig. 1 shows him both when his general average is changing, as after week 16, or when it is fluctuating, as in weeks 1–4.

Linear Drift

Another kind of chart on which single estimates appear is a sales control chart on which the money value of weekly sales is plotted. This frequently has the complication of a seasonal effect like that in Fig. 2. The standard method of calculating the residual variation about the trend lines involves a regression analysis and analysis of variance, and for the sales data of Fig. 2 gives a standard deviation of 4·6. The standard deviation of 4·9 estimated from the mean successive difference is a good approximation to this figure since it measures the variations from one point to the next irrespective of their average level, and only

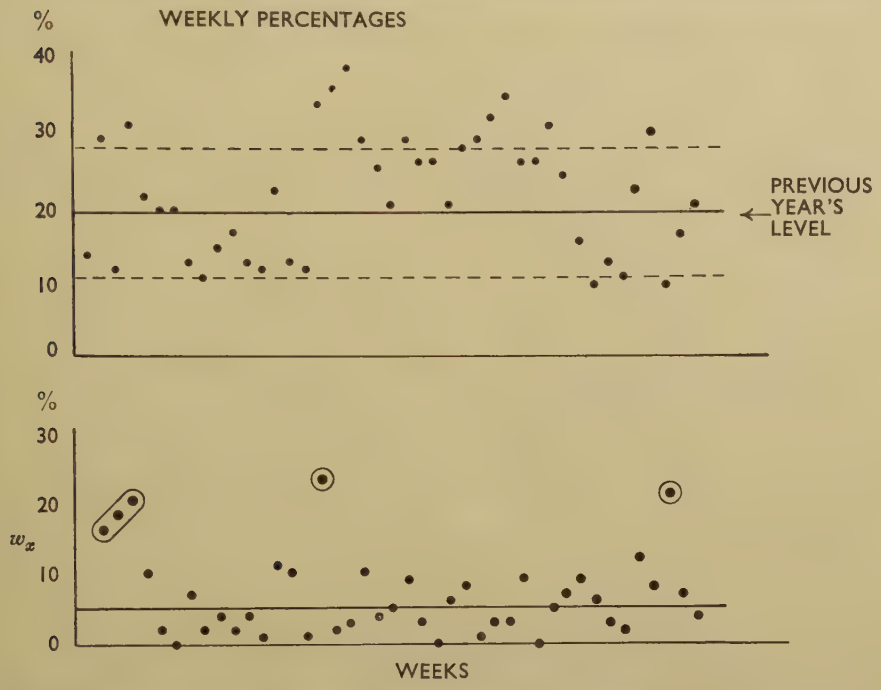


FIG. 1. Weekly percentage of grade 2 crockery for the home market. The upper chart gives the weekly percentages and the lower one the successive differences, w_x. The mean successive difference, excluding ringed values, is 183/37 = 4·94, and the estimated standard deviation is 4·94/1·13 = 4·4 per cent.

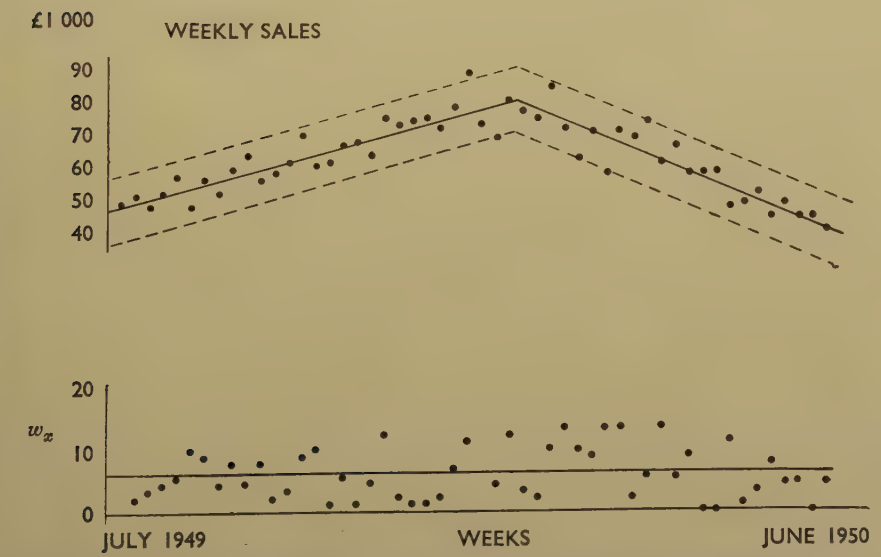


FIG. 2. Weekly sales of electric lamps (upper chart) and successive differences (lower chart). The mean successive difference is 2·81/51 = 5·52 and the estimated standard deviation 5·52/1·13 = 4·9.

a very small part of the change in average level is included in each of the successive differences.

Once the seasonal pattern is known the trend lines with the 'two-sigma' limits can be drawn for each new year and the weekly sales plotted as they become available. If points gradually fall below the bottom limit, then a sales drive can be instituted to try to check the decline. If a 'target' value of five per cent. more than the previous year's sales is set, the position of the trend lines can be raised correspondingly.

Cyclic Effect

In many kinds of technical and economic data in which only single estimates are available there are cycles of the kind shown in Fig. 3. There are several empirical methods of drawing a smoothed curve through the points, but rarely is any estimate of the residual variation made. Estimation of the residual variation by formal methods, though reasonably practicable for the previous straight-line example, is always arduous when the movement is not linear, even if the form of the cycle is known, but becomes extremely difficult if it is not. Again the mean successive difference can be used to calculate the short-term variability, enabling control chart limits to be set about the smoothed curve. It will be noted that for the temperatures of Fig. 3 no formally calculated standard deviation is given for comparison with the estimated standard deviation of 3·4.

If good dimensional control of blown glass articles is to be maintained it is essential that the temperature of the furnace in which the glass is melted should not vary by too much. The width of the limits in Fig. 3 is based on what is normally achieved, and future readings compared with the limits will show whether the variability is as good as usual and whether the pattern is maintaining itself.

Control Charts showing External Variability

Under the normal procedure for measurement control charts, small samples of fixed size are taken at regular intervals and the sample averages and ranges plotted on a pair of charts. Control limits can then be set based on the within-sample (internal) range such that, if the process is in strict control, a high proportion of the sample averages will lie within them.

However, it is frequently found in industrial practice that there is a random and inevitable variability between sample averages which is greater than would be expected from the variability to be found within the samples, and this causes the averages frequently to fall outside the usual control limits based on the within-sample variation. Since we do not wish to base any action on this 'external' variability σ_e, which is no less random than the 'internal' variability σ_i, limits are used which embrace both components of variation. In order to watch the separate effects of them both and to estimate their magnitude a third

chart is added which records the successive differences between *sample averages*, $w_{\bar{x}}$. Fig. 4 shows the three charts for some electric lamps. The upper chart for averages has limits based on

$$\sqrt{(\sigma_e^2 + \sigma_i^2/n)}$$

where $n = 4$ is the number in each sample; the chart detects changes in level which are greater than the combined 'internal' and 'external'

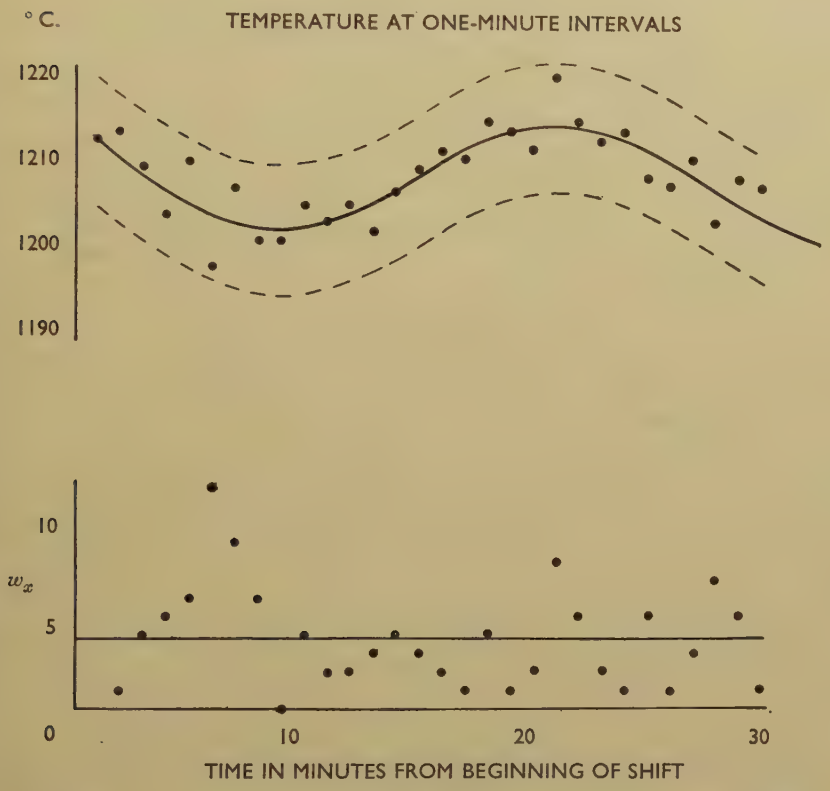


FIG. 3. Temperature in a regenerating glass furnace and (lower chart) successive differences. The mean successive difference is $111/29 = 3{\cdot}83$ and the estimated standard deviation $3{\cdot}83/1{\cdot}13 = 3{\cdot}4$.

variations. The centre chart for differences between successive averages gives separate sample estimates of

$$1{\cdot}13\sqrt{(\sigma_e^2 + \sigma_i^2/n)}$$

and the lower chart for ranges within samples gives sample estimates of $\sigma_i d_n$, where d_n is the factor relating the mean range and the standard deviation in samples of n.

If the magnitude of the external component, σ_e, is explicitly required, σ_i must first be estimated by dividing the mean range $\bar{w}_n$ by the appropriate factor d_n in the usual way; this is then subtracted from the standard error of the averages, $\sqrt{(\sigma_e^2 + \sigma_i^2/n)}$, estimated from the

successive difference chart by dividing the mean successive difference $\bar{w}_{\bar{x}}$ by $d_2 = 1{\cdot}13$. Then

$$\sigma_e^2 = \left(\frac{\bar{w}_{\bar{x}}}{1{\cdot}13}\right)^2 - \frac{1}{n}\left(\frac{\bar{w}_n}{d_n}\right)^2$$

The *total variability* between individuals selected at random, provided no change in level has occurred, is then given by

$$\sigma_t^2 = \sigma_e^2 + \sigma_i^2$$

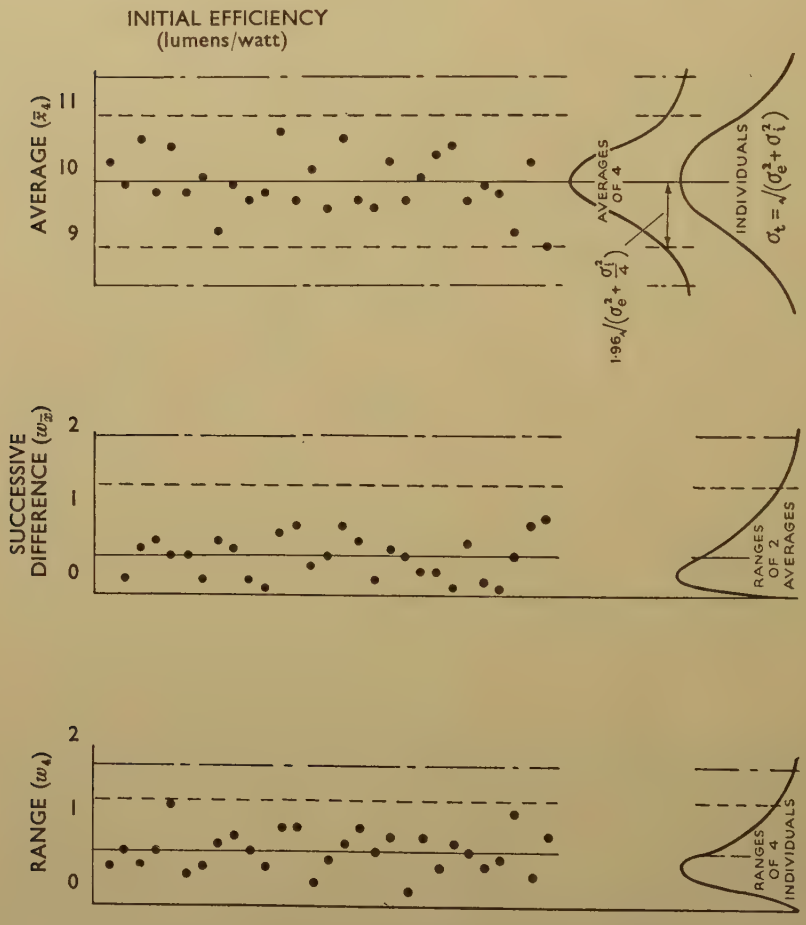


FIG. 4. Control charts for samples of four lamps, allowing for external as well as internal variability.

Example. Fig. 4 shows the type of chart that is used as evidence of conformity to a specification for the granting of a British Standards Institution quality mark. It is known that there is a variability between batches of tungsten lamp filaments in addition to the variation within each batch (many thousands of lamps are made from a batch), and each source of variability is closely controlled.

In Fig. 4 the averages, successive differences between averages, and ranges are shown for 15 consecutive batch samples. The grand mean of 9·85 is used as the estimate of the process average. From the range chart the mean range in samples of four, $\bar{w}_4$, is found to be 0·7, from which the internal standard deviation is estimated as $\sigma_i = 0{\cdot}7/2{\cdot}06 = 0{\cdot}34$. From the successive difference chart the mean successive

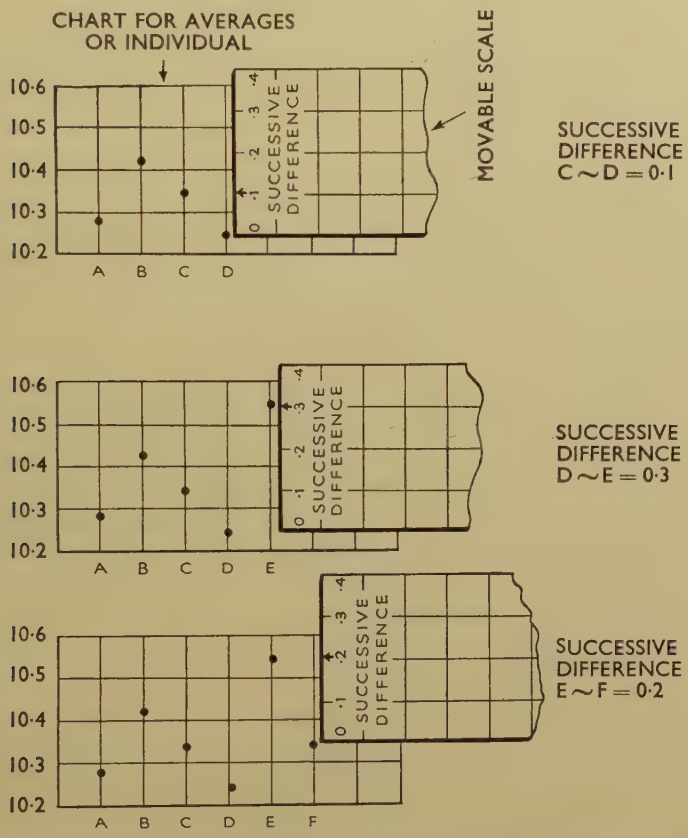


FIG. 5. Diagrams illustrating a quick method of reading successive differences from a chart of averages or individual readings.

difference is 0·5, and the standard error of the averages is estimated as $0{\cdot}5/1{\cdot}13 = 0{\cdot}44$. Hence $\sigma_e^2 = 0{\cdot}44^2 - \frac{1}{4}0{\cdot}34^2 = 0{\cdot}165$ and $\sigma_e = 0{\cdot}41$. The total variability between individuals when no change in level has occurred is then $\sigma_t = \sqrt{(\sigma_i^2 + \sigma_e^2)} = \sqrt{(0{\cdot}34^2 + 0{\cdot}41^2)} = 0{\cdot}53$.

Control limits can be set on the successive difference chart as for ranges in twos, although, because of the positive correlation between one difference and the next arising from the use of each observation twice, the interpretation of these limits is slightly modified, as Dr Hartley shows in the Appendix. This chart can then be used to keep a check on the external variability. It will also be influenced slightly

by the internal variability, but, if the external variability is large enough to be significant, at least three-quarters of the variability shown on the chart will be due to the external component. The range chart will, as usual, keep a running record of the internal variability, and the average chart of the general level.

The above estimates of the internal and external variation could again have been obtained by simple analysis of variance provided that no drifts or other changes in level had taken place. We feel, however, that the method outlined may be useful to those who are unfamiliar with the formal methods or who have no calculating aids available.

Change in Level

The method can be extended to situations in which a big change in average level has occurred. If the change in level is associated with a very large successive difference, this value should be excluded from the calculation. If the text-book method of calculating the standard deviation is used under these circumstances, an unreal figure is obtained which includes both the within-level variability and the effect of the change in level.

The method of estimating variability outlined in this article is very much quicker than more formal methods without any sensible loss in accuracy for chart purposes. It can be made even more speedy if a scale equivalent to that on the vertical axis of the chart is drawn on a scrap of graph paper, extending from zero to the maximum successive difference. If this scale is then turned parallel to the vertical axis and the zero is set against the lower of a pair of successive readings as in Fig. 5, the difference can be easily read off. This is an invaluable aid if past records involving thousands of readings are being studied.

APPENDIX: SOME THEORETICAL ASPECTS

HERMAN O. HARTLEY
University College, London

The purpose of this note is to place on record and appraise certain theoretical properties of the method of estimating variability described in the foregoing paper.

The notation is in terms of the example given in Table I. The n ($= 20$) original readings 10·3, . . ., 10·4 are denoted by x_1, . . ., x_n, their 19 successive differences 0·8, . . ., 1·0 by $w_1 = |x_2 - x_1|$, . . ., $w_{19} = |x_{20} - x_{19}|$, and the average of these $(0{\cdot}8 + \ldots + 1{\cdot}0)/19 = 0{\cdot}45$ by $\bar{w}$. If the variation in the readings is purely a normal distribution, which may comprise both internal and external components of variation, its standard deviation σ can be estimated by $\hat{s} = \bar{w}/1{\cdot}128$, where the divisor 1·128 is the familiar mean range d_2 ($= 2/\sqrt{\pi}$) in samples of two.

This estimate $\acute{s}$ is computationally simpler than the familiar standard deviation* $s = \sqrt{\{\Sigma(x - \bar{x})^2/(n-1)\}}$; it is also unbiased, i.e. its mean value in repeated samples of readings is σ. However, $\acute{s}$ is less reliable than s. It can be shown[1] that the variance of s is given by

$$\{4\pi/3 + 2\sqrt{3} - 6 + (4 - 2\sqrt{3} - \tfrac{1}{3}\pi)/(n-1)\}\sigma^2/2(n-1)$$
$$= \{1{\cdot}653 - 0{\cdot}5113/(n-1)\}\sigma^2/2(n-1) \quad \ldots.(1)$$

or for moderately large samples approximately by $\sigma^2 1{\cdot}653/2n$. Comparing this with the familiar large sample variance of s, which is $\sigma^2/2n$, we see that the variance of $\acute{s}$ is 1·65 times that of s, and if it is computed from $n = 100$ observations it has the same precision as a sample standard deviation s computed from 61 observations.

In spite of this disappointing loss in precision there may well be situations in which readings are 'cheap' and plentiful and the method involving the computation of $\acute{s}$ has practical advantages. Nevertheless I would stress here that the outstanding merits of $\acute{s}$ lie in its comparative freedom from bias in situations of non-random variation which are discussed below. To emphasise the point that in *purely random variation* the use of $\acute{s}$ requires caution I shall now show that, in fact, there are other short-cut estimates of σ which may be preferred to $\acute{s}$.

Suppose we had split the above sample of 20 readings into four consecutive sub-samples of five consecutive observations and computed the range for each sub-sample, we could have estimated σ as follows:

sub-sample ranges .. 0·8, 1·5, 0·5, 1·0
mean range .. .. $3{\cdot}8/4 = 0{\cdot}95$
range estimate of σ = mean range/d_5 = $0{\cdot}95/2{\cdot}326 = 0{\cdot}41$

This estimate is close to that obtained above (0·40) and is, in fact, of higher precision. To illustrate the comparative accuracy of range estimates I give below for a sample of 20 the variances of the estimates of standard deviation from mean range in:

(*a*) 2 sub-samples of 10, $k = 2$, $m = 10$ variance = $0{\cdot}033\,5\sigma^2$
(*b*) 4 sub-samples of 5, $k = 4$, $m = 5$ variance = $0{\cdot}034\,5\sigma^2$
(*c*) 5 sub-samples of 4, $k = 5$, $m = 4$ variance = $0{\cdot}036\,5\sigma^2$
(*d*) 10 sub-samples of 2, $k = 10$, $m = 2$ variance = $0{\cdot}057\,1\sigma^2$

The variance of $\acute{s}$ for a sample size $n = 20$ evaluated from equation (1) is $0{\cdot}042\,8\sigma^2$ and is seen to be larger than those of the above estimates, except that of estimate (*d*) based on the mean of $k = 10$ ranges of $m = 2$ readings, which only utilises every second successive difference between an odd and even observation falling into the same subsample. A general result[2] states that in estimating σ from the mean range of k

* The mean value of s^2 is equal to σ^2 exactly, but the mean value of s is equal to σ only approximately as the sample size increases.

samples of size m it is most advantageous to choose for m a convenient sample size near $m = 8$ such that $km = n$.

Now it will be seen that in the practical examples of the preceding paper the encountered variation is certainly not random, but that the readings are affected by a slowly moving linear 'drift' as in Fig. 2 or by a cyclic variation as in Fig. 3. Clearly, if we must estimate the standard deviation σ representing uncontrollable random variation (as we surely must in quality control chart technique) we must exclude such systematic trends from our measure of variation. This can, of course, be done by estimating σ from the deviations of the readings from the fitted trend or other regression law, but such a procedure would require an exact knowledge of the type of trend law underlying the data. If regression analysis is not used and σ is estimated from the overall standard deviation, s, the systematic drift or other systematic movement would seriously increase this estimate of σ and consequently all control limits based on such an estimate would be too wide and no real control could be established. On the other hand, if σ is estimated from $\hat{s}$ the bias will usually be negligible. Formulae for the respective biases in the two estimators can be worked.* They depend on the type of trend in the readings. Let the ith reading x_i be composed as $x_i = \theta_i + z_i$, where θ_i is the ordinate of the trend law and z_i a superimposed random error variable with standard deviation σ. Then it can be shown that

$$\text{mean of } s^2 = \sigma^2\{1 + \Sigma(\theta_i - \bar{\theta})^2/(n-1)\sigma^2\} \quad \ldots.(2)$$

$$\text{mean of } \hat{s} = \sigma\{1 + \Sigma(\Delta\theta_i)^2/4(n-1)\sigma^2\} \quad \ldots.(3)$$

where $\Delta\theta_i = \theta_{i+1} - \theta_i$ and terms in $(\Delta\theta_i/\sigma)^4$, etc., have been ignored in equation (3). If all $\theta_i = \text{constant} = \bar{\theta}$, i.e. if there is no systematic trend, both mean values are equal to σ^2. If, however, the θ_i differ it is clear that the bias in $\hat{s}$ will be much smaller than that in s since it depends only on the $\Delta\theta_i$, i.e. the differences between *consecutive* θ_i, which will be small for 'gradually changing' trend laws. To illustrate the respective bias of s and $\hat{s}$ we consider as an example a linear trend $\theta_i = (0{\cdot}1\sigma)i$ where $i = 1, 2, \ldots, 25 = n$. This means that there is a systematic upward step from reading to reading amounting to one-tenth of the σ of the superimposed error variable. Substituting in formulae (2) and (3), we find that the bias in $\hat{s}$ is $\frac{1}{4}$ per cent. while that in s^2 is 54 per cent., so that the bias in s is $100\sqrt{(1{\cdot}54 - 1)} = 24$ per cent. Clearly, where the possibility of such systematic movements cannot be excluded we do well to employ Miss Keen's and Mr Page's estimator $\hat{s}$ as a safeguard against bias.

A final remark concerns the interpretation of the second control chart of Fig. 4, on which successive differences are plotted and judged against the ordinary control limits for the range in samples of two. In

* A more comprehensive study of the estimator $\hat{s}$ will shortly be published by Mr A. R. Kamat of University College, London, to whom I am indebted for formulae (2) and (3).

an *ordinary* control chart for the range in samples of two we would expect the upper inner limit to be exceeded in 2·5 per cent. of cases when the process is under control. In inspecting such a chart on which say l ranges are plotted one would therefore bear in mind the following two properties: (*a*) that the expected number of ranges exceeding the inner upper limit is $0{\cdot}025l$, and (*b*) that the chance of at least one range exceeding the inner upper limit is $1 - (1 - 0{\cdot}025)^l$, or approximately $0{\cdot}025l$. Now, as the authors remark, there is a considerable positive correlation between the successive differences, and the question arises: How does this correlation affect statements (*a*) and (*b*)? It is clear that (*a*) is completely unaffected; but the chance (*b*) is considerably smaller. I hope shortly to publish the precise values for the chances (*b*). The reason for this apparent anomaly is the 'clustering' of the successive differences because of their positive correlation.

REFERENCES

[1] Arley, N., & Hald, A. (1950). *Matematisk Tidsskrift*, **86**.
[2] Grubbs, F. E., & Weaver, L. C. (1947). *J. Amer. Statist. Ass.*, **42**, 224.

SHORT-TERM FLUCTUATIONS IN PETROLEUM CONSUMPTION

JOHN DOUGLAS BUTLER
Shell Petroleum Co. Ltd.

In this article Mr Butler not only shows how he has evolved a method of estimating the trend in petroleum consumption but also expresses the uncertainty in the resulting forecasts by giving limits within which short-term fluctuations may be expected to lie.

Introduction

This article describes an attempt made in early 1952 to predict the future extent of short-term fluctuations in the consumption of petroleum products in an area comprising the world outside the USA, the USSR, and the Soviet orbit.

For the oil industry a knowledge of the probable limits of such fluctuations is becoming increasingly important. Before the war gasoline was the most important single product in terms of volume, and financially was the product which yielded the highest net return. Since 1938, however, the consumption of refined products has not only increased very rapidly, but has changed considerably in pattern. Whereas in 1938 the heavier products known as the black oils (gas oil, diesel fuel, fuel oil) were 50 per cent. of the consumption in the area studied, by 1951 they had risen to 60 per cent. of the total. In the same period the relative importance of gasoline had fallen from 36 to 26 per cent. of the consumption. Between 1938 and 1951 heavy fuel oils alone increased from 33 to 40 per cent. of the total consumption of refined products. The post-war oil industry is, therefore, supplying a vastly expanded demand predominantly composed of the black oils.

A pilot investigation into energy consumption in the OEEC countries had shown that whereas there was but a small correspondence between gasoline consumption and the index of industrial production, black oil consumption tended to reflect changes in the index though these changes were superimposed upon a rapid growth trend which was probably the result of the extension of black oil consumption in those fields of use to which it was peculiarly suited. It was therefore felt desirable to attempt to measure the probable limits of short-term fluctuations in petroleum consumption both because of the possibility of contractions in consumption during an overall recession, but also because of the need to evaluate the probable upper limits to consumption should the momentum of the post-war expansion prove not yet to be exhausted.

Method

Estimates were already available from an annual survey of supply and demand. This survey is based on information derived from questionnaires sent out to representatives in all the major and nearly all the minor countries concerned. The completed questionnaires provide, *inter alia*, estimates of the consumption of individual products for the current year and the next five years. The questionnaire is accompanied by instructions to the compilers to ignore the possibility of shortages of supplies. It was felt at the time this study was begun that, in view of the buoyancy of petroleum consumption in the post-war period, the estimates of consumption in future years might possibly reflect the unconscious assumption that the recent trend in the level of business activity would continue in the period under review. At other times the estimates may well be influenced by 'slump' thoughts.

The problem was thus twofold: in the first instance it was necessary to project a long-term trend for petroleum consumption over a period of five years in the future, which is relatively short-term in relation to the base period, and secondly it was necessary to establish the probable short-term fluctuations around this trend. It was also desirable to be able to relate the survey estimates to the limits of fluctuation so established. As a result of the pilot investigation of energy consumption in the OEEC countries, referred to above, it appeared that a primary requirement was a continuous series that could be used as an indicator of the general level of economic activity. It was decided therefore that a period covering at least twenty-five years should be used for ascertaining the historical trends in economic activity and petroleum consumption.

No series was available, either for economic activity, or for industrial production extending from 1920 to 1951, and the construction of an index of industrial production for this period proved impracticable. It was thought that an index of total energy consumption would be a sufficiently representative series for this purpose, however, since the pilot investigation had established that there was a high degree of correlation between industrial production and energy consumption in the industrial countries of Western Europe, the coefficient of correlation between the logarithms of the two series being 0·97.

The known tendency for gross production to increase at a slightly higher rate than energy consumption is attributed to increased efficiency of utilisation; its effects were therefore partly eliminated by weighting the consumption of the various forms of energy not only according to their thermal content, but also in accordance with their efficiency in use; for most of the period the preponderance of coal makes this adjustment unimportant.

Consumption figures for coal, petroleum, natural gas, and hydro-electricity were therefore prepared. In the earlier years these had in some cases to be based on production net of export/import balances and in all cases stock changes were ignored. From these figures an index was prepared and was used as an indicator of the course of general economic

activity as shown in Fig. 1. The index has been calculated by using the average of the whole period as a base. Though the value of this procedure would perhaps be questionable if it were to be used for the comparison of levels of the index in years in the future, it facilitates comparison of two series over the same historical period.

The next stage was to prepare an index of petroleum consumption in terms of crude oil. A comparison of the two series indicated that though petroleum consumption has been influenced by changes in the general

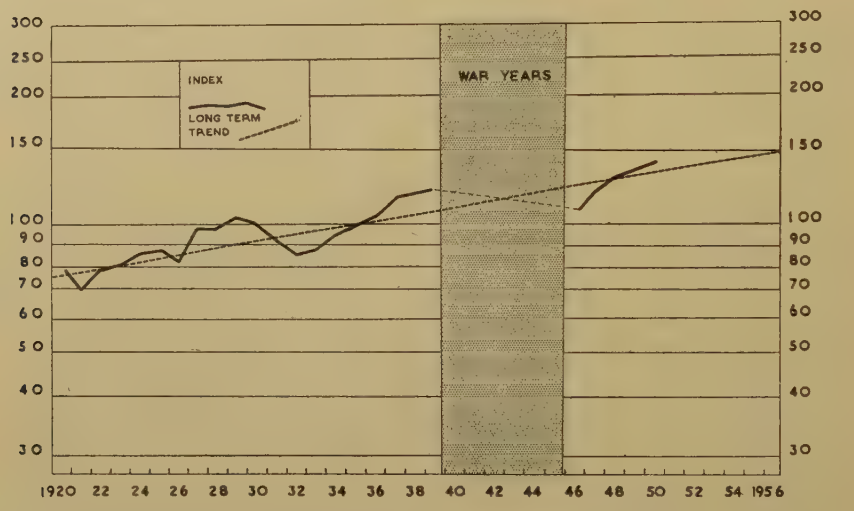


Fig. 1. The course of general economic activity for the World excluding USA and the Soviet sphere. 1920–50 = 100.

level of economic activity, its inherent rate of growth has been a more important factor. This confirmed the conclusions of the above-mentioned pilot investigation into energy consumption in Western Europe and it indicated the necessity for assessing separately the importance of the inherent rate of growth. The data of historical petroleum consumption were therefore 'deflated' by dividing the petroleum index by the index of general economic activity provided by the index of total energy consumption. This gave a fairly smooth curve which seemed as though it might quite well be represented by a gompertz function, and this was accordingly calculated giving the equation:

$$\log Y_c = 2{\cdot}437\ 70 - 0{\cdot}811\ 31(0{\cdot}960\ 51)^x$$

where Y_c is the calculated trend value and x is the year with 1920 as origin. The trend has been drawn with the smoothed index in Fig. 2. As the chart shows, the data are well represented by the trend-line, though there is a tendency for the calculated values to fall below the actual values in the last three post-war years.

There arises at this juncture a point of theoretical importance. It may well be that the future will reveal that in the post-war period there is the point of inflection of a new gompertz cycle. This could be the inherent pattern of the growth of petroleum consumption or could be the result of the war. There is no doubt that the needs of post-war reconstruction exercised a strong stimulating effect on petroleum consumption and there is nothing essentially illogical in the supposition that as a result of the war a completely new growth cycle commenced.

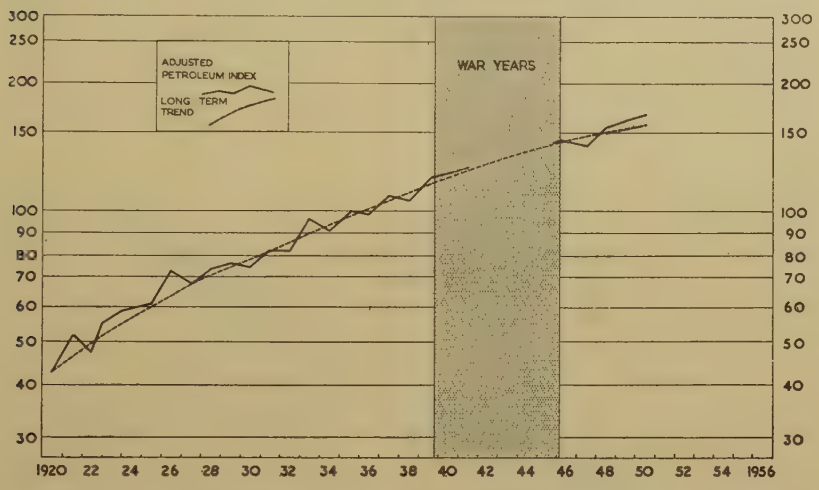


FIG. 2. Petroleum consumption with the influence of general economic changes removed. 1920–50 = 100.

Evidence of the growth of consumption in what are now known as the 'under-developed' countries lends some support to this theory.

Practically, however, this consideration makes little difference since the projection to be made is over a period of only five years, though it must be kept in mind when the more distant future of petroleum consumption is contemplated. In order, however, to avoid the possibility of understating the post-war growth-trend, the trend was phased forward four years to give a closer fit to the data. The ultimate results of the study were the basis for this decision, since it was more important to present the survey figures in relation to the most optimistic assumptions about the long term trend than in comparison with the minimum level of consumption possible in a recession. The standard error of the curve, which was not significantly affected, was 1·05 per cent.

A composite function was then calculated for the unsmoothed petroleum index by combining the individual values of the logarithms of the exponential trend fitted to the index of general economic activity with those for the gompertz curve fitted to the smoothed index. The standard error of this composite trend was approximately 8 per cent.

Results

The method was then ready for application to the problem of estimating the limits of short-term fluctuations in the level of petroleum consumption. A projection of the long-term economic trend combined with the trend in the 'deflated' petroleum index was made to give a forecast of the trend in petroleum consumption over the next five years. This was not in itself of great direct use, but by combining the standard errors of the two curves a band of probable fluctuation was set out; it was considered that the 'two-sigma' limits gave a sufficiently good idea of the probable limits to future fluctuations. The index of petroleum consumption was then converted into actual consumption figures in

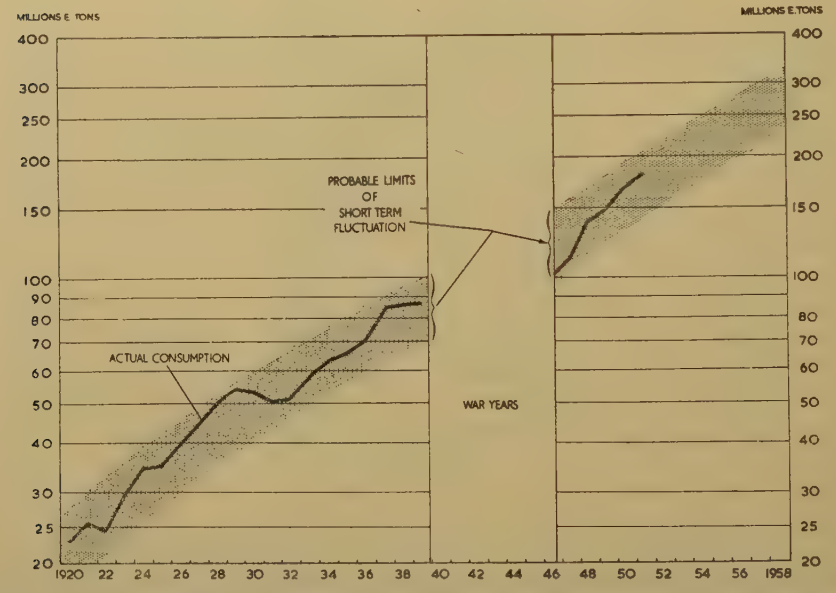


FIG. 3. The consumption of refined product for the World excluding USA and the Soviet sphere. 1920–50 = 100.

terms of crude oil, as were the trend values, and a deduction was made for refinery use and loss (i.e. that proportion of the crude oil consumed which is used in fuelling the refinery production processes) to reduce the resultant figures into terms of refined product consumption. The results of the study were then presented on the lines of Fig. 3.

The results of the survey figures, which were compiled later, fell within the probable limits of fluctuations so established. The chart made it easy to explain how the survey estimates might possibly be regarded in the light of current economic trends and more particularly gave an indication of the degree of flexibility which the supply side of the industry might be required to display in the course of the next few years.

It may be considered that a gompertz curve could be fitted directly to the consumption data. This, however, would not be satisfactory; for not only would it produce a large unexplained variance, but also one of the aims of this study was to indicate to management, in quantitative terms, to what extent the future growth of demand was determined by factors outside the influence of the industry. Thus subjective judgment of the general course of business activity could be related back to the consequent behaviour that should be expected from petroleum consumption. The method used has proved useful in this connection.

The technique described in this study is simple, both because of the need to convey its implication to the lay reader and also because paucity of statistical material relating to world oil consumption makes more detailed assessments hazardous; but I feel that it may represent a step along the road towards the objective of estimating the short-term limits within which the petroleum industry must plan to operate. The evaluation is of course short-term in every way; it is undoubtedly insufficiently refined to permit forecasts for periods over much more than five years. The actual limits established are also rather broad, but it is good to face the uncertainty of any projection since there is already too strong a tendency to make forecasts apparently down to improbably close limits, thus encouraging delusions of accuracy.

A STATISTICAL APPROACH TO A PROBLEM OF CHEMICAL MANUFACTURE

ANDREW MUIR *Fisons Limited* and
FRANK DOWNTON *Imperial College, London*

When a chemical reaction takes place in a laboratory the quantities of the chemicals can be controlled accurately to the equivalent weights and the whole of the 'input' chemicals can be used up in the products. When the same reaction occurs on an industrial scale it may be impracticable to control the quantities accurately and usually there is an excess of one or other 'input' chemical with a consequent wastage and loss in yield. Mr Muir and Mr Downton deal with the statistics of this situation and show how to calculate the most economical conditions.

Introduction

We shall be concerned with irreversible reactions which take place in solution and in which one of the two end-products is either precipitated or is easily removed. The results may, however, be extended to other types of reactions, the only limitation being that there are two combining chemicals or two 'inputs' as we shall call them. For example, the action of hydrochloric acid on sodium silicate in aqueous solution to obtain the insoluble silicic acid is a reaction of this type:

$$Na_4SiO_4 + 4HCl \rightarrow H_4SiO_4 + 4NaCl$$

We shall assume that an excess of one of the inputs over the other will not affect the yield of the required end-product and that the excess input will be entirely wasted. This assumption is clearly only an approximation and its nearness to the true state of affairs will depend upon the solubilities of the reacting chemicals. For an end-product of very low solubility, however, this assumption may be justified. The situation which arises when the solubility of the end-product is not sufficiently low is also discussed.

With this assumption, suppose we represent the reacting chemicals by

$$A + B \rightarrow C + D$$

where A and B are the 'inputs' and C is the required end-product. Let the equivalent weight ratio be $a : b : c : d$ respectively. Then

(i) a of A + b of B will give c of C.
(ii) a' of A + b of B will give c of C if $a' > a$.
(iii) a of A + b' of B will give c of C if $b' > b$.

It will be shown that on an industrial scale using a units of A and b units of B a lower yield of C than c units is to be expected because of inaccuracies and variations in a and b. It is then suggested that although it is impossible to obtain c units of C it is possible to set up a criterion of efficiency of the process and to adjust the 'inputs' so that this efficiency is maximised.

The Problem of Depressed Mean Yield

It is a simple matter in the laboratory to combine a units of A with b units of B and get almost exactly c units of C. On the industrial scale, however, the input chemicals may be measured into the reacting tank by orifice or by some other method less precise than the laboratory balance, and at best we can only achieve an average of a units of A and b units of B. Intuitively we would therefore expect to obtain an average yield of C lower than c units. In Appendix I we show that by making the assumption already mentioned and by assuming that the inputs are normally distributed about means $a(1 + \alpha)$ and $b(1 + \beta)$ with standard deviations σ_1 and σ_2 respectively, a formula may be obtained to calculate the expected mean yield of C. If we put $V^2 = \sigma_1^2/a^2 + \sigma_2^2/b^2$ this expected yield is given by:

$$c(1 + \beta) - c(\beta - \alpha)\frac{1}{\sqrt{(2\pi)}}\int_{-\infty}^{(\beta-\alpha)/V} e^{-t^2/2}dt - cV\frac{1}{\sqrt{(2\pi)}}e^{-(\alpha-\beta)^2/2V^2} \quad \ldots . (1)$$

Both $\frac{1}{\sqrt{(2\pi)}}\int_{-\infty}^{U} e^{-t^2/2}dt$ and $\frac{1}{\sqrt{(2\pi)}}e^{-U^2/2}$ are fully tabulated in most statistical textbooks or in books of statistical tables, so that the expected yield may be easily calculated.

It may be seen that when $\alpha = \beta = 0$, i.e. when we are in fact averaging the intended input weights, our expected yield becomes

$$c\left(1 - \frac{V}{\sqrt{(2\pi)}}\right) \quad \ldots . (2)$$

Thus we can never achieve the required output c units of C unless $\sigma_1 = \sigma_2 = 0$ or, alternatively, in expression (1) $\beta = 0$ and $\alpha = \infty$. It naturally follows that we must investigate the problem of maximum efficiency of this process when σ_1 and σ_2 are not zero.

Example. In a certain process 149 units of A reacted in solution with 144 units of B. The theoretical yield of the precipitate C was 220 units. A and B were measured into the reacting tank by calibrated orifices and it was found that after 60 batches the average yield was only 212 units. It was required to know whether this could be explained by errors in the orifice measurements.

It was found that orifice A averaged 144 units and that orifice B also averaged 144 units, and that $\sigma_1^2/a^2 = 0{\cdot}0016$ and $\sigma_2^2/b^2 = 0{\cdot}0012$.

Thus, $a(1 + \alpha) = 144$, and since $a = 149$, $\alpha = -0{\cdot}033$. Similarly, $\beta = 0$, $V = 0{\cdot}0529$, and $(\beta - \alpha)/V = 0{\cdot}62$. Then from equation (1) we see that the expected mean yield is

$$220 - 220 \times 0{\cdot}033 \times \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{0{\cdot}62} e^{-t^2/2} dt - 220 \times 0{\cdot}053 \times \frac{1}{\sqrt{(2\pi)}} e^{-0{\cdot}62^2/2}$$

and this is approximately 211. Clearly the low yield could be explained by the orifice errors.

Maximum Efficiency and Maximum Expected Profit

It is clear that the actual yield of C can be expressed as a percentage of the intended yield of C. Thus if an average of a units of A and b units of B are put into the reacting tank and a and b are equivalent weights, the percentage yield of C may be deduced from equation (2) as $100\{1 - V/\sqrt{(2\pi)}\}$. If, however, there is a bias in one or both of the inputs (i.e. α or $\beta \neq 0$) it is difficult to know what the intended yield of C is, since it could be taken as that resulting from weights equivalent to $a(1 + \alpha)$ or $b(1 + \beta)$ or to a and b.

Since there is no unique measure of the efficiency in this sense, some other criterion must be introduced. If all other things are equal this criterion will clearly be expressed in terms of money, or at least in the relative prices of the chemicals concerned. Our criterion of 'maximum efficiency' may then be regarded as the 'maximum expected profit.' We define 'expected profit' as the average gross profit, and 'gross profit' as the selling price of the product less the cost of the materials used. Later we introduce the term 'net profit,' which is the selling price less the cost of the chemicals and the plant running costs. In Appendix II we show how a consideration of this 'maximum expected profit' gives rise to an equation in α and β which may be reduced to a usable form in the following way.

It is assumed that $\alpha = 0$; this may be done without loss of generality, since β may have positive and negative values. In considering the cost of chemicals it is necessary to distinguish between the cost to use and to waste, since it may be too costly to reclaim excess materials and they may have to be treated before disposal. Let the cost for chemicals be as follows:

chemical A per unit weight, λ_1 to use and λ_2 to waste, and
chemical B per unit weight, μ_1 to use and μ_2 to waste.

All other costs are assumed constant for each batch.

Let the selling prices be:

product C per unit weight, η,
product D per unit weight, ν.

Then the average value of the gross profit is shown in Appendix II to be:

$$\bar{G} = c\eta + d\nu - a\lambda_1 - b\mu_1 - b\mu_2\beta \int_{-\infty}^{\beta/V} \frac{1}{\sqrt{(2\pi)}} e^{-t^2/2} dt$$

$$+ \beta(a\lambda_2 + c\eta + d\nu - a\lambda_1 - b\mu_1) \int_{-\infty}^{-\beta/V} \frac{1}{\sqrt{(2\pi)}} e^{-t^2/2} dt$$

$$- \left\{ V(c\eta + d\nu - a\lambda_1 - b\mu_1) + \frac{1}{V}\left(\frac{a\lambda_2\sigma_2^2}{b^2} + \frac{b\mu_2\sigma_1^2}{a^2} \right) \right\} \frac{1}{\sqrt{(2\pi)}} e^{-\beta^2/2V^2} \quad \ldots . (3)$$

This is at a maximum when β has a value that satisfies equation (4):

$$\left\{ 1 + \frac{d\nu}{c\eta} - \frac{b(\mu_1 - \mu_2)}{c\eta} - \frac{a(\lambda_1 - \lambda_2)}{c\eta} \right\} \int_{-\infty}^{-\beta/V} \frac{1}{\sqrt{(2\pi)}} e^{-t^2/2} dt$$

$$- \frac{b\mu_2}{c\eta} - \frac{\beta}{V^3}\left\{ \frac{\sigma_1^2}{a^2} \cdot \frac{a\lambda_2}{c\eta} + \frac{\sigma_2^2}{b^2} \cdot \frac{b\mu_2}{c\eta} \right\} \frac{1}{\sqrt{(2\pi)}} e^{-\beta^2/2V^2} = 0 \quad \ldots . (4)$$

Equation (4) cannot be solved for β directly. The best method is to try a few values of β, evaluate the left-hand side of (4), and then by numerical or graphical interpolation find the value that satisfies the equation.

Example. In the same process as in the previous example it was required to know what bias should be set up for the most economical running of the plant.

Cost of chemical A was £16 per ton and it required no neutralising if wasted, so that $\lambda_1 = \lambda_2 = £16$ per ton. Chemical B was a by-product of another process and was highly corrosive, so that if it passed through the system it cost £10 per ton to be neutralised. Thus, $\mu_1 = 0$ and $\mu_2 = £10$ per ton. The selling price of chemical C was £50 per ton and that of D was zero, so that $\eta = £50$ per ton and $\nu = 0$.

The equation of process is

$$A + B \rightarrow C + D$$

and the equivalent weights are

$$\left.\begin{array}{rcl} 149 + 144 & \rightarrow & 220 + 73 \\ a + b & \rightarrow & c + d \end{array}\right\}$$

Hence $a/c = 0{\cdot}6773$, $b/c = 0{\cdot}6545$, $d/c = 0{\cdot}3318$, and as before, $V = 0{\cdot}0529$. If we substitute in equation (4) and put $\gamma = \beta/0{\cdot}0529$, we have

$$\int_{-\infty}^{-\gamma} \frac{1}{\sqrt{(2\pi)}} e^{-t^2/2} dt - 0{\cdot}1157 - 0{\cdot}1591 \frac{\gamma}{\sqrt{(2\pi)}} e^{-\gamma^2/2} = 0$$

The solution of this is $\gamma = 1{\cdot}019$, whence $\beta = 0{\cdot}054$.

Now if we consider only the costs and prices of the materials and calculate as a measure of the efficiency of our process the average gross profit per unit value of the product C, i.e. $\bar{G}/c\eta$, we find from equation (3):

$$\text{for } \beta = 0, \qquad \bar{G}/c\eta = 0{\cdot}7632$$

and $$\text{for } \beta = 0{\cdot}054, \quad \bar{G}/c\eta = 0{\cdot}7738$$

The increase in the gross profit due to using the optimum excess of chemical B as compared with using, on the average, the chemically equivalent quantity is 1·4 per cent.

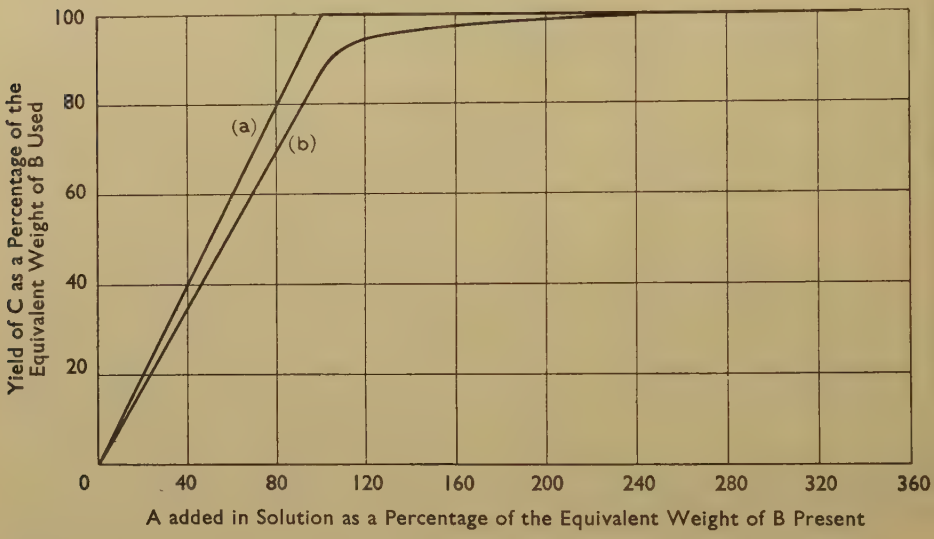


FIG. 1.

The running costs of the plant, which are assumed to be constant, enter into the selling price of the products. If these amount to $0{\cdot}7c\eta$, a not unreasonable amount, and they are subtracted from the gross profit, we have:

$$\text{for } \beta = 0, \qquad \text{expected net profit}/c\eta = 0{\cdot}0632$$

and $$\text{for } \beta = 0{\cdot}057, \quad \text{expected net profit}/c\eta = 0{\cdot}0738$$

The increase measured on this basis is as much as 16·8 per cent. and is equivalent to a profit increase of 11s. per ton.

Solubility

The effect of solubility on the production of a precipitated end-product can be represented by curves of the type shown in Fig. 1. The curve (*a*) is that which results if no solubility effect is present and is the curve on which the results so far obtained is based. If the solubility of C is about 1 in 150 a curve similar to (*b*) will be obtained. Since,

however, such curves are based on only one of the 'inputs' it does not easily show what additional waste is introduced by the solubility effect.

This additional waste is represented by the vertical difference between curves (*a*) and (*b*) and is expressed as a percentage of the equivalent weight of B used. Fig. 2 shows this additional waste, which is due to solubility, expressed as a percentage of that input which is not in excess. An example should make this clear.

Suppose we have 0·8*a* of A reacting with 1·0*b* of B. The theoretical yield of C is 0·8*c*, which shows a wastage of 20 per cent. of the equivalent weight of B present, and no wastage when expressed as a percentage of

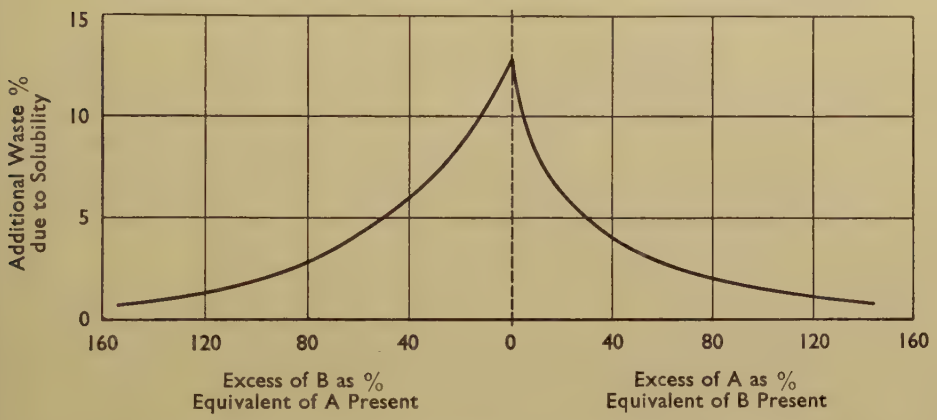


FIG. 2. The additional waste is expressed as a percentage of the quantity of A or B present, whichever is smaller.

the equivalent weight of A present. Fig. 1 shows that the actual yield is 0·7*c*, which shows an additional wastage of 10 per cent. of the equivalent of B present or 8 per cent. of the equivalent of A present. If now we look at Fig. 2 we see that for a 25 per cent. excess of B over A, the additional wastage expressed as a percentage of A present is 8 per cent.

It may be possible by assuming an exponential or some other form for the two curves in Fig. 2 to introduce the effect of solubility into the calculations already performed. Certainly the effect of solubility will not make the use of the calculated bias less efficient than using no bias at all, although the calculated bias will not give maximum efficiency.

Conclusion

The results we have obtained seem simple enough and agree with the intuitive suggestion that it is sensible to use an excess of the cheaper of the two input chemicals to use up variations in the dearer of the inputs. They go further, however, in that they estimate how large this excess should be. The results are accurate for only a very low solubility

of the required end-product, yet even if the solubility is not low enough to be neglected it will still be more profitable to use an excess of the cheaper of the two inputs.

In order to apply the results of this paper it is necessary to estimate the standard deviations of the inputs. It should not be difficult to do this from a number of observations, and the estimate is accurate enough if its error is not greater than 1 per cent. of the weight of the corresponding material.

APPENDIX I. DERIVATION OF EQUATION FOR YIELD

We consider the chemical reaction

$$\mathrm{A} + \mathrm{B} \rightarrow \mathrm{C}\ (+\ \mathrm{D})$$

and adopt the notation given in the article.

Let the actual weights of chemicals for a given batch be x of A, y of B, and z of C, and let these be values of variates X, $\mathcal{Y}$, and $\mathcal{Z}$. The probability distributions of X and $\mathcal{Y}$ have already been stated. Then if $\Pr\{\ \}$ denotes the probability of the results signified in the brackets:

$$\Pr\{z \leqslant \mathcal{Z} \leqslant z + dz\} = \Pr\{x \leqslant X \leqslant x + dx\}\Pr\{\mathcal{Y} \geqslant y\} + \Pr\{y \leqslant \mathcal{Y} \leqslant y + dy\}\Pr\{X \geqslant x\}$$

where $x/a = y/b = z/c$.

Let the probability distribution function of z be $\psi(z)$, and write

$$\frac{z - c - \alpha c}{c\sigma_1/a} = u \quad \text{and} \quad \frac{z - c - \beta c}{c\sigma_2/b} = v$$

$$\text{Then } \psi(z) = \frac{1}{2\pi} \cdot \frac{1}{c\sigma_1/a}\, e^{-u^2/2} \int_v^\infty e^{-t^2/2} dt + \frac{1}{2\pi} \cdot \frac{1}{c\sigma_2/b}\, e^{-v^2/2} \int_u^\infty e^{-t^2/2} dt$$

The expected yield is

$$\mathrm{E}(z) = \int_{-\infty}^{\infty} z\psi(z) dz$$

and when evaluated this leads to equation (1).

APPENDIX II. DERIVATION OF EQUATIONS FOR GROSS PROFIT

We use the same notation as before and write the normal probability distribution functions of x and y respectively as $\phi_1(x)$ and $\phi_2(y)$. Then if the variate G is the gross profit and g is a particular value we have:

$$\text{for } x/a < y/b,\ g = \frac{cx}{a}\eta + \frac{dx}{a}\nu - x\lambda_1 - \frac{bx}{a}\mu_1 - \left(y - \frac{bx}{a}\right)\mu_2$$

$$= k_1 x + \mu_2 y;$$

for $x/a > y/b$, $g = \frac{cy}{b}\eta + \frac{dy}{b}\nu - \frac{ay}{b}\lambda_1 - y\mu_1 - \left(x - \frac{ay}{b}\right)\lambda_2$

$$= k_2 y - \lambda_2 x,$$

where $k_1 = \frac{c\eta}{a} + \frac{d\nu}{a} - \lambda_1 - \frac{b}{a}(\mu_1 - \mu_2)$

and $k_2 = \frac{c\eta}{b} + \frac{d\nu}{b} - \frac{a}{b}(\lambda_1 - \lambda_2) - \mu_1.$

Then

$$\Pr\{g < G \leqslant g + dg\} = \int_{x=-\infty}^{x=\infty} \phi_1(x)\phi_2(\nu)\Pr\{x/a \leqslant y/b\}dxdy$$

$$+ \int_{y=-\infty}^{y=\infty} \phi_1(x)\phi_2(y)\Pr\{x/a > y/b\}dxdy$$

$$= \frac{1}{\mu_2}\int_{x=-\infty}^{x=\infty}\phi_1(x)\phi_2\left(\frac{k_1}{\mu_2}x - \frac{g}{\mu_2}\right)\int_{y=bx/a}^{y=\infty}\phi_2(y)dydxdg$$

$$+ \frac{1}{\lambda_2}\int_{y=-\infty}^{y=\infty}\phi_1\left(\frac{k_2}{\lambda_2}y - \frac{g}{\lambda_2}\right)\phi_2(y)\int_{x=ay/b}^{x=\infty}\phi_1(x)dxdydg$$

Let the probability distribution function of g be $\theta(g)$.

Then:

$\theta(g) = \Pr\{g < G \leqslant g + dg\}/dg$

$$= \phi\left\{\frac{g + \mu_2 b(1+\beta) - k_1 a(1+\alpha)}{\sqrt{(k_1^2\sigma_1^2 + \mu_2^2\sigma_2^2)}}\right\}$$

$$\Phi\left\{-\frac{ab(\alpha-\beta)\sqrt{(k_1^2\sigma_1^2 + \mu_2^2\sigma_2^2)}}{\sigma_2\sqrt{[\mu_2^2(a^2\sigma_2^2 + b^2\sigma_1^2) + a^2k_1^2\sigma_1^2]}}\right.$$

$$\left. - \frac{k_1\sigma_1^2 b[g + \mu_2 b(1+\beta) - k_1 a(1+\alpha)]}{\sigma_2\sqrt{(k_1^2\sigma_1^2 + \mu_2^2\sigma_2^2)}\sqrt{[\mu_2^2(a^2\sigma_2^2 + b^2\sigma_1^2) + a^2k_1^2\sigma_1^2]}}\right\}$$

$$+ \phi\left\{\frac{g + \lambda_2 a(1+\alpha) - k_2 b(1+\beta)}{\sqrt{(\lambda_2^2\sigma_1^2 + k_2^2\sigma_2^2)}}\right\}$$

$$\Phi\left\{\frac{ab(\alpha-\beta)\sqrt{(\lambda_2^2\sigma_1^2 + k_2^2\sigma_2^2)}}{\sigma_1\sqrt{[\lambda_2^2(a^2\sigma_2^2 + b^2\sigma_1^2) + b^2k_2^2\sigma_2^2]}}\right.$$

$$\left. - \frac{k_2\sigma_2^2 a[g + \lambda_2 a(1+\alpha) - k_2 b(1+\beta)]}{\sigma_1\sqrt{(\lambda_2^2\sigma_1^2 + k_2^2\sigma_2^2)}\sqrt{[\lambda_2^2(a^2\sigma_2^2 + b^2\sigma_1^2) + b^2k_2^2\sigma_2^2]}}\right\}$$

where $\phi\left\{\frac{u-\delta}{\epsilon}\right\} = \frac{1}{\epsilon\sqrt{(2\pi)}} e^{-(u-\delta)^2/2\epsilon^2}$

and $\Phi\{u\} = \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{u} e^{-t^2/2} dt$

The mean gross profit $\bar{G}$ is

$$\mathrm{E}(G) = \int_{-\infty}^{\infty} g\theta(g)dg$$

and this gives equation (3) when $\alpha = 0$. In order to maximise $\mathrm{E}(G)$ we evaluate $\partial \mathrm{E}(G)/\partial\beta = 0$ to give equation (4).

A COMPARISON OF SERIAL NUMBER DIGIT SAMPLING WITH SYSTEMATIC AND RANDOM SAMPLING

ORVILLE O. PRICE and ALFRED J. McKEON
Research Division, US Veterans Administration

This is an interesting example of the application of standard statistical methods to show that the sampling technique described gives effectively random samples in a particular case.

Introduction

One of the basic sampling techniques used by Research Division, USVA, to study the US veteran population may be called 'Serial Number Digit Sampling,' and consists of selecting veterans with service serial numbers ending in one or more randomly selected digits. It has generally been assumed that this method gives results about the same as systematic or random selection. The purpose of this study is to determine empirically whether certain samples of World War II veterans so selected can be treated as random samples in estimating sampling errors.

Sample Size

Serial number digit sampling applied to a single series of numbers consisting of consecutive integers would yield a 10 per cent. sample if a single digit (0–9) were used, a 1 per cent. sample if two digits (00–99) were used, and so on. The samples so obtained would really be systematic, consisting of every 10th, 100th, etc., item in the population. However, there are about twenty series of service serial numbers of World War II veterans (male and female officers and enlisted of the Army, Air Force, Navy, Marine Corps, and Coast Guard), none of which necessarily consists of consecutive integers, and the corresponding samples do not represent systematic selection in the usual sense. The numbers in each series were assigned by blocks representing specified areas of military command, and since the numbers for each area were assigned to personnel consecutively in order of original entrance into the armed forces, proportional representation would be expected from each area.

In order to test, for World War II veterans, the randomness of the method of selection described, 14 435 punched cards recording separations from the US Armed Forces were obtained, constituting a 0·1 per cent. sample of all World War II veterans with service serial numbers ending in 267. The results outlined below apply only to sub-samples from these 14 435 veterans. This file of punched cards was tabulated on all groups of 4th and 5th digits from the end of the serial number,

and the frequencies are shown in Table I. For example, the number of veterans with serial numbers ending in 00267 is 166; in 0267 is 1 512.

TABLE I

Distribution of US World War II veterans with service serial numbers ending in 267, by fourth and fifth digits from last

Fifth Digit from Last	Fourth Digit from Last										Total
	0	1	2	3	4	5	6	7	8	9	
0	166	165	159	141	151	161	152	154	161	175	1 585
1	165	153	150	154	152	152	160	159	158	149	1 552
2	156	146	156	141	154	149	158	149	154	147	1 510
3	154	161	148	149	150	152	150	141	152	149	1 506
4	158	152	145	143	142	151	141	150	136	152	1 470
5	143	124	139	139	143	144	146	137	142	140	1 397
6	143	149	148	155	143	138	141	137	143	146	1 443
7	147	139	144	130	144	147	143	146	137	133	1 410
8	147	135	128	139	132	127	128	135	129	127	1 327
9	133	117	124	125	123	121	130	117	123	122	1 235
Total	1 512	1 441	1 441	1 416	1 434	1 442	1 449	1 425	1 435	1 440	14 435

It appears that selection from this group by use of the fourth digit yields a sub-sample of about 0·01 per cent. of the original population. However, selection by use of a pair of fourth and fifth digits (00–99) gives a sub-sample slightly more than 0·001 per cent. if the fifth digit is small (0, 1, etc.) and slightly less than 0·001 per cent. if the fifth digit is large (9, 8, etc.). This is probably due in part to some of the twenty series mentioned above not being complete for as many as five significant digits.

The percentage of non-whites in each of the 100 sub-samples is shown in Table II. Since the sub-samples vary in size (Table I), the χ^2 test,

$$\chi^2 = \sum_i [(f_i - pn_i)^2 / pqn_i]$$

with $p = 0{\cdot}0824$, $q = 0{\cdot}9176$, and f_i and n_i the numbers of non-whites and veterans respectively in the ith sub-sample, was applied to test the hypothesis that the sub-samples could be regarded as random samples of size n_i. The value of χ^2 was found to be 89·66; $n' = 99$; $P = 0{\cdot}74$, which is consistent with the hypothesis.

Estimation of a Proportion from a Sample

If each sub-sample of Table II is considered to be a random sample

of 144 veterans, the standard error of the estimated proportion is given approximately by

$$\sigma_p = \sqrt{(pq/n)} = \sqrt{(0{\cdot}824 \times 0{\cdot}9176/144)} = 0{\cdot}0229 = 2{\cdot}29 \text{ per cent.}$$

TABLE II

Distribution of sub-sample estimates of percentage non-white based on selection by last five digits of service serial number

Fifth Digit from Last	Fourth Digit from Last									
	0	1	2	3	4	5	6	7	8	9
0	6·02	11·52	7·55	8·51	7·95	11·18	10·53	8·44	6·83	5·14
1	9·09	7·19	9·33	6·49	7·24	7·24	8·75	10·06	10·13	11·41
2	13·46	11·64	8·97	9·93	9·74	8·05	11·39	10·74	8·44	6·80
3	9·09	7·45	4·73	6·71	8·67	8·55	8·00	7·09	7·89	8·72
4	7·59	9·87	6·90	6·99	9·15	5·96	9·22	10·00	11·03	6·58
5	6·99	4·84	10·07	12·23	8·39	8·33	13·70	8·03	8·45	8·57
6	9·79	8·72	10·81	6·45	7·69	10·14	8·51	10·95	7·69	10·27
7	4·76	4·32	6·94	10·00	4·86	3·40	4·20	8·22	6·57	6·77
8	3·40	8·15	7·03	9·35	8·33	6·30	3·91	9·63	9·30	5·51
9	9·02	10·26	7·26	2·40	9·76	5·79	10·77	6·84	8·13	9·02

As in the systematic sampling design, the theory of the serial number digit design follows directly from the simplest case of cluster sampling where only one cluster is selected as the sample and there is no sub-sampling within clusters.[1] In the systematic design each cluster consists of every kth item in the population, and there are k clusters of the same size. In the serial number digit design each cluster consists of all items in the population with serial numbers ending in the same digit or digits, and there are 10, 100, etc., clusters which are usually not the same size.

In the simple cluster sampling design where the population of N elements is divided into k clusters of n elements each, and one of the clusters is taken as the sample, each of the clusters has the probability $1/k$ of being the one selected and $N = kn$. For this design the variance of a sample proportion is given by the formula

$$\sigma_p^2 = [1 + (n - 1)\rho]pq/n$$

where p is the population proportion and ρ is the intra-class correlation coefficient and is given by the equation

$$\rho = \left[\frac{\sum_{i=1}^{k}(p_i - p)^2}{k} - \frac{\sum_{i=1}^{k} p_i q_i}{k(n-1)}\right] \Big/ pq$$

where p_i is the proportion in the ith cluster and k is the number of clusters in the population.[2]

In clusters formed by serial number digits n is not usually the same for all clusters but a good approximation to the variance is given by substituting $\bar{n}$ for n in the above formula, where $\bar{n}$ is the average size, provided no cluster is more than say 25 per cent. larger or smaller than the average.

If each sub-sample of Table II is considered to be a cluster sample with $\bar{n} = 144$ and $p = 0{\cdot}0824$, use of the formula just given and the values of p_i from Table II give $\sigma_p' = 0{\cdot}0217 = 2{\cdot}17$ per cent., which is not significantly different from the 2·29 per cent. estimated by random theory. In this particular case ρ is nearly zero, and therefore the sampling variability is about the same as for a random sample.

For large random samples, sample proportions are approximately normally distributed about the population proportion with a standard deviation equal to the standard error of the sample proportion. To test this hypothesis for the sub-samples of Table II a normal curve was fitted, using $p = 0{\cdot}0824$ and $\sigma_p = 0{\cdot}0217$. The observed and theoretical frequencies are shown in Table III. The theoretical and observed

TABLE III

Frequency distribution of sub-sample estimates of percentage non-white

Percentage Non-white	Frequency of Sub-samples		Percentage Non-white	Frequency	
	Observed	Theoretical		Observed	Theoretical
Under 4·70	7	5·2	8·70–9·69	14	16·5
4·70–5·69	5	6·8	9·70–10·69	14	12·2
5·70–6·69	8	12·0	10·70 and over	13	12·9
6·70–7·69	20	16·2	Total ..	100	100·0
7·70–8·69	19	18·2			

frequencies are compared by calculating $\chi^2 = \Sigma(f_o - f)^2/f$, where f_o is the observed frequency, f the corresponding theoretical frequency, and Σ the summation over all frequency groups. In this case $\chi^2 = 4{\cdot}01$, there are 5 degrees of freedom, and $P = 0{\cdot}55$, which is the probability of a fit as bad or worse because of chance.

Estimation of a Mean from a Sample

The means (average age 30th June 1950) of the 100 sub-samples are shown in Table IV. The grand average age is 31·47 years and the standard deviation 6·78 years. If we consider each sub-sample to be a random sample of 144 veterans, the standard error of the sub-sample mean is given approximately by

$$\sigma_{\bar{x}} = \sigma/\sqrt{n} = 6{\cdot}78/12 = 0{\cdot}567.$$

The unweighted standard deviation of the 100 sub-sample means is 0·531 years, which is also an estimate of $\sigma_{\bar{x}}$ and is not significantly different from the estimate of 0·567. The unweighted mean of the sub-sample means is 31·44 years.

TABLE IV

Sample mean ages of World War II veterans, based on selection by last five digits of service serial number (deviations from 30·00 years)

Fifth Digit from Last	Fourth Digit from Last									
	0	1	2	3	4	5	6	7	8	9
0	2·24	2·60	1·76	1·20	2·40	2·61	1·07	1·84	2·22	1·00
1	2·15	1·31	0·88	2·11	1·43	1·20	0·79	0·96	0·84	0·80
2	2·16	1·52	1·67	2·17	1·97	2·33	2·40	2·12	1·75	1·62
3	0·99	1·77	0·99	1·14	0·99	1·63	1·95	1·15	1·49	1·36
4	0·89	1·50	1·82	2·39	1·97	1·34	1·22	2·43	1·61	1·24
5	1·89	0·71	1·31	1·52	0·22	0·87	1·57	0·69	0·44	1·27
6	1·45	1·59	0·86	1·20	0·58	1·05	1·52	0·99	1·52	1·03
7	1·64	1·15	0·81	1·11	0·95	0·50	1·81	1·61	0·85	1·31
8	1·41	1·22	1·86	1·73	1·89	2·04	0·89	1·74	1·47	1·93
9	1·23	1·28	1·60	0·77	0·86	2·14	1·95	1·03	1·11	0·59

The frequency distribution of the means of Table IV has been compared with the theoretical distribution—a normal distribution with a mean of 31·44 and a standard deviation of 0·531, and the two do not differ significantly.

Conclusion

As far as these limited data are concerned, samples of World War II veterans selected by the use of service serial number digits may be treated as random in estimating sampling errors.

REFERENCES

[1] Madow, Lillian H. (1946). 'Systematic Sampling and its Relation to Other Sampling Designs.' *J. Amer. Statist. Assoc.*, **41**, 204.

[2] Hansen, Morris H., & Hurwitz, William N. (1943). 'On the Theory of Sampling from Finite Populations.' *Ann. Math. Statist.*, **14**, 333.

STOCHASTIC PROCESSES OR THE STATISTICS OF CHANGE*

MAURICE S. BARTLETT
University of Manchester

The term 'stochastic processes' is used to denote statistical phenomena that develop in time and the theoretical models that arise in their treatment, and as these are encountered in many fields, this article has a wide practical interest. Professor Bartlett indicates how stochastic processes arise in physics and communication engineering, industry, economics, biology, and medicine. As an example he describes the construction of a mock epidemic series simulating successive measles outbreaks in a boarding school.

What are Stochastic Processes?

A few years ago not many people in this country knew what was meant by a stochastic process; today the situation is perhaps no different for the general public, though professional statisticians are becoming more familiar with the phrase. Like other new phrases or words, its use 'catches on,' and it gradually spreads through the community of statisticians or other receptive agents much in the way an actual infection will spread. Such a process is itself a *stochastic* process, by which is meant that it does not proceed according to any immutable law but is at least partly dependent on random and chance factors. We therefore call it a *random* or *stochastic* process, usually preferring the second adjective because *random* might convey the idea that every stochastic process appeared purely haphazard (like the emissions from a radioactive substance or the so-called Brownian motion of small dust particles on the surface of a liquid), whereas in many stochastic processes, such as the spread of epidemics or the growth of populations, any random fluctuations may be apparently eliminated by the large statistical groups involved, so that the development of the process appears comparatively smooth and even predetermined.

Anyone previously unfamiliar with the idea of a stochastic process will by now be beginning to see what is meant. He may, even if his mathematics is a relic from his schooldays, remember the distinction in mechanics between statics and dynamics. If he is an economist he will know that much of classical economic theory is erected on the same kind of static or equilibrium structure as classical statistical mechanics, and will know that modern economic theorists, like their physicist colleagues, are busy trying to formulate their theories to

* This article is based on a paper given at the Joint Conference of the Royal and Manchester Statistical Societies held in Manchester in September 1952.

represent a little more closely our dynamic changing world. If he is a statistician he will know that mathematical statisticians long ago began to study the statistical populations and frequency distributions arising in nature and how far they may be represented by theoretical models which assist in their interpretation. *The theory of stochastic processes is, roughly speaking, concerned with the corresponding wider theory of the statistics of change.*

Interpreting the subject in this way, we can be either excited by its generality or disappointed by its lack of novelty. Of course, stochastic processes have always been there in nature, and the industrial statistician studying his control charts or the commercial statistician his firm's fluctuating sales figures will not automatically solve his problems by calling them by a new name. However, while the statistician should never be hidebound by the standard techniques available to him, there is a limit to the extent to which even the best statistician can make *ad hoc* improvisations on current methods, and there is no doubt that recent systematic study of the theory of stochastic processes has greatly broadened his possible approach to actual statistical problems. I will cite two or three examples of how the older 'static' outlook tended to be a barrier to improved technique until it was broken down.

The first was in the statistical analysis of time-series.* The 'static' procedure of considering a given sample of *independent* observations had of course been adapted as far as possible to the study of time-series even in classical methods, in which the stochastic process was represented by a trend or a harmonic curve to which independent random fluctuations were supposed added; but even this assumption proved too narrow to cover many cases met with in practice. In particular, the impossibility of such an assumption always being feasible became apparent from the case of continuous time-records. In this case the assumption of independence implied unlimited statistical information if the discrete observations taken over a fixed period of time were increased indefinitely by reducing the interval between successive observations. Historically, the first attacks on this important statistical problem using a more general approach were made independently by the Russian mathematician and econometrician E. Slutsky[23] and by the English statistician Udny Yule[25] in the year 1927.

The second example was in the practice of industrial sampling. The 'single sampling' schemes were supplemented by 'double sampling,' 'inverse sampling,' and finally by the 'sequential sampling' methods, whose theory was mainly developed during the last world war by the American mathematical statistician Abraham Wald.[7] With these sequential methods the new feature is the *continuation* of sampling until enough information has been acquired for a decision to be taken with a specified risk. Sequential sampling is thus much less 'static' than the classical practice of taking a sample of predetermined size, and its

* This general heading strictly includes such topics as control charts, especially if the successive entries in the latter turn out to be correlated.

distributional theory (e.g. determining the average size of sample required in any application) is essentially one falling within the general field of stochastic process theory and, in particular, is related to the 'random walks' and diffusion processes referred to in the next section.

As a third example consider the manner in which discrete frequency distributions arise—in particular the well-known Poisson distribution for small numbers. In text-books this is usually derived from the binomial distribution, but a more direct and in many cases a more natural way is to obtain it as the fundamental distribution associated with events occurring randomly and independently of each other in time, such as the emission of alpha-particles by a radioactive substance or, in suitable cases, the occurrence of accidents to a particular individual or at a particular locality. The theoretical derivation is comparatively simple. If the total number of events occurring in the time t is $\mathcal{N}(t)$, then the distribution of $\mathcal{N}(t)$ can be specified by its 'probability-generating function'

$$G(z; t) \equiv p_0(t) + p_1(t)z + p_2(t)z^2 + \ldots,$$

in which the coefficient $p_r(t)$ of z^r is the probability that $\mathcal{N}(t) = r$ after a time t. We suppose that in a small time-interval δt the chance of one extra event occurring is $a\ \delta t$ (for simplicity we assume a constant in time, though this is not essential), and the chance of none, $1 - a\ \delta t$. Then

$$G(z; t + \delta t) = a\ \delta t\, zG(z; t) + (1 - a\ \delta t)G(z; t)$$

or

$$dG/dt = a(z - 1)G,$$

whence

$$G = e^{at(z-1)}$$

if $\mathcal{N}(0) = 0$. This is the probability-generating function of the Poisson law $p_r(t) = e^{-m}m^r/r!$, with a mean $m = at$. The theory of stochastic processes thus gives the Poisson distribution a basic role in statistical theory not less than that of any other distribution.

Of course, if the events (e.g. accidents) are not independent the distribution may be modified. An important case is that where the chance of an event in the small time-interval δt is not constant, but depends on the number of events that have already occurred, being of the form $[a + b\mathcal{N}(t)]\ \delta t$. This 'contagion' hypothesis may be shown by an extension of the above method to lead to the 'negative binomial distribution,' a result first established as long ago as 1914 by A. G. McKendrick,[4] who was interested in its medical applications. Another way in which this same distribution can arise was discovered in 1920 by Greenwood and Yule, who were investigating the numbers of accidents experienced by a group of munition workers. They found that if these workers were variable in their proneness to accidents then the frequency distribution of the numbers of accidents per individual, obtained from the statistics of the whole group, may be of the 'negative binomial' type (see, for example, Lundberg[3]). If we wish to discriminate between the first hypothesis, that any individual may suffer more

accidents than the average because initial accidents contracted by bad luck render him more liable to others, and the second hypothesis, that one individual will differ from another in his accident proneness right from the start, it is necessary to analyse the accidents per individual over more than one time-period. This has recently been done, for example, on statistics collected for South African shunters.[1]

The fact that more than one causal mechanism can generate the same statistical distribution is an obvious warning to the statistician who is hoping to learn something of the way an observed distribution may have arisen. Before embarking on such a task he should ideally be familiar with *all* the theoretical possibilities. Even so, without some further limitation of the possible hypotheses, the extent to which he can unravel data presented to him may be severely limited. This difficult but vital problem of what the statistician is entitled to ask before undertaking a statistical analysis, particularly in connection with stochastic processes, is returned to again later.

The Monte Carlo Method

While systematic study of stochastic processes is recent, it is evident that in various guises they have appeared since the concepts of probability and chance were first formulated. It is in fact remarkable how the early mathematicians in their attacks on probability problems raised by gamblers included studies of game sequences closely related to many modern stochastic process problems, such as sequential analysis or the use of artificial stochastic processes to solve differential equations and other theoretical problems (the so-called 'Monte Carlo method'). For example, in 1657 the famous Dutch mathematician C. Huyghens propounded the following problem (quoted from a paper by Professor G. A. Barnard[17]). 'A and B each take twelve counters and play with three dice on this condition, that if eleven is thrown, A gives a counter to B, and if fourteen is thrown, B gives a counter to A; and he wins the game who first obtains all the counters. Show that A's chance is to B's as 244 140 625 is to 282 429 536 481.' The mathematical equation for this problem is readily set up, for if the chances of obtaining fourteen or eleven at any trial are as $p : q$ (actually 15 : 27 in this case), the probability $P(x)$ of A winning when he has x counters must satisfy the 'difference equation'

$$P(x) = \frac{p}{p+q} P(x+1) + \frac{q}{p+q} P(x-1), \quad (0 < x < 24) \quad \ldots\ldots(1)$$

and also $P(0) = 0$, $P(24) = 1$. The relevant solution is

$$P(x) = [(q/p)^x - 1] / [(q/p)^{24} - 1] \quad \ldots\ldots(2)$$

or in particular

$$P(12) = 1 / [(q/p)^{12} + 1] \quad \ldots\ldots(3)$$

agreeing with Huyghens's answer.

The interesting point about this is that, if the theoretical solution of the problem, which has been formulated in mathematical terms in equation (1), had not been precisely known, an approximate solution could be obtained not by direct numerical methods, but by repeated simulation of the gambling problem. As a simple illustration, one hundred repetitions were made, all of which resulted in A losing, consistently with the true value of P being as low as 0·000 864. To expedite these repetitions they were made for convenience with the aid of four-figure random numbers rather than of dice. To obtain odds of 15 : 27 we may classify any random number into one of the two groups 0000–3570 and 3571–9999, giving the practically equivalent odds 3571 : 6429. Notice how the artificial games, which simulate real ones, can give us at the same time all possible information we may wish to know. For example, we know the number of trials required before a game is terminated, and so accumulate information on the statistical distribution of the 'length' of a game. This problem can also be solved theoretically, but the mathematical solution is quite complicated.

This 'artificial sampling' or 'Monte Carlo method' is well known to statisticians, so much so that tables of random numbers are a familiar item in their libraries. In recent years, however, it has also been seriously considered by mathematicians as an aid to the solution of differential or other mathematical equations (see, for example, the USA publication on the 'Monte Carlo method'[10]). Thus the above gambling problem is an example of what is called a 'random walk' process in which each 'step' (in this case the transfer of a counter) occurs independently of previous steps; if we further consider the individual steps of this 'walk' to be small compared with the total distances to be traversed (in the gambling problem the number of counters originally held must be comparatively large), it may be shown that the density f of 'paths' when a large number of repetitions of the process is envisaged satisfies the partial differential equation

$$\partial f/\partial t + a\,\partial f/\partial x = b\,\partial^2 f/\partial x^2 \qquad \ldots.(4)$$

where x is the net distance traversed in the 'time' t (the number of steps), $a = (p - q)/(p + q)$, and $b = 2pq/(p + q)^2$. This equation is well known as the 'equation of diffusion' in physics, and conversely, if we met this equation directly and wished to obtain a solution of it by the Monte Carlo method, we could choose an appropriate $p : q$ (altering if necessary the scale of t and hence of a and b), and proceed as already indicated. Of course, we should be unlikely in this rather simple case to use this method in practice unless the boundary conditions were more complicated or we required to accumulate a lot of information simultaneously about the underlying process. However, to illustrate the connection with equation (4), the frequency distribution of the *lengths* of the 'games' in the hundred repetitions already referred to is compared in Fig. 1 with the theoretical distribution from (4) of the time required to reach a boundary at distance 12 units from the starting

point. The agreement, with so few stages as 12 to reach the boundary in the artificial games, is surprisingly good.

It might be added that the reason a stochastic process may so often be found corresponding to equations arising in physics is the obvious one that the equation has really arisen in the first place, as in the case of equation (4), from a stochastic process occurring in nature.

Stochastic Processes in Physics and Communication Engineering

The reader will, however, appreciate that this more fundamental role of stochastic processes in physical problems cannot be adequately

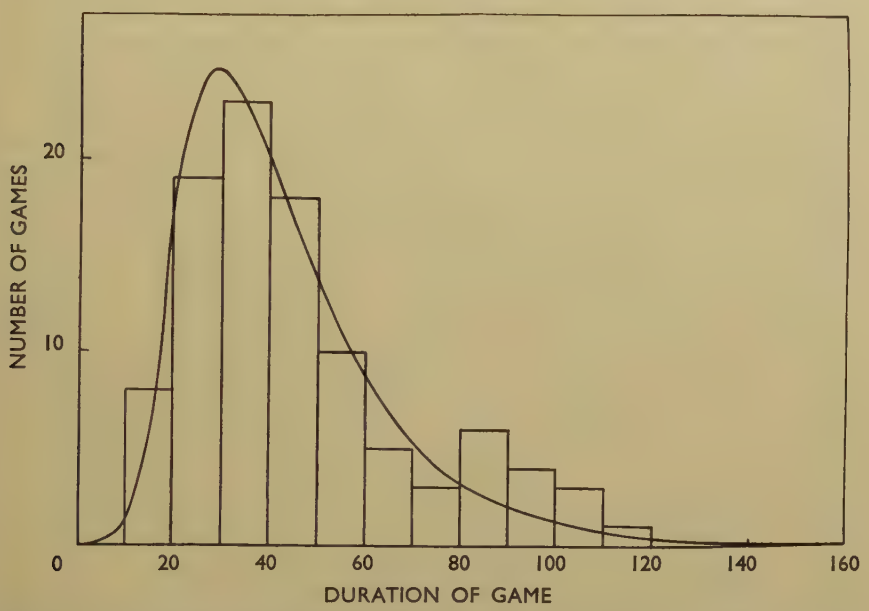


FIG. 1. Frequency distribution of the duration of play in 100 'games,' yielding an approximate solution of a diffusion problem whose correct solution is given by the continuous curve.

indicated here. It must suffice to remind him that the explicit use of the theory of stochastic processes for such physical phenomena as the showers of particles created by cosmic rays or other 'chain reactions,' as Brownian motion or other 'noise' phenomena, or as turbulence in gases and liquids, is merely one indication of the increasing necessity with all physical processes to allow fully for the role which chance and statistical concepts play in them.

Two topics may perhaps be singled out as of particular interest in industrial physics and communication engineering. The first is concerned with the stray disturbances in electrical and other delicate apparatus and has already been referred to above by its usual title, 'noise.' The problem of correcting and filtering a long train of signals to reduce the effect of noise to a minimum is one that comes under the general theory of what are called 'stationary' time-series, and has

been so treated by Norbert Wiener, the American mathematician.[15] By a *stationary* time-series is meant a stochastic process in time in which the variable is fluctuating or oscillating but not otherwise changing as time goes on (an example is shown in Fig. 2). From the point of view of the communication engineer a continuing sequence of messages or signals may often be regarded as a stationary time-series, whether or not they are affected by external random disturbances. For example, in a long passage in English the way in which the various letters or even words happen to follow each other has a definite and constant statistical structure which can be studied. If someone sends such a passage by teleprinter or communicates it verbally by telephone, the resulting electrical signals will also constitute a stationary time-series.

The second topic is concerned with the *communication* and *coding* aspect of these series of messages or signals. It is clear that while the sequence of electrical signals should, apart from the effect of disturbances, represent the original sequences of messages, there is a considerable choice in how the representation is made; and one method may be better than another. Here again the concept of stationary time-series is used in the construction of a general theory of communication, in which are studied and made precise such questions as: what is the maximum rate of information that may be passed along a given channel, or equivalently, how 'big' has a channel to be to pass a required rate of information? To give some idea, to those familiar with the technical jargon of the electrical engineer, of the kind of results that can be reached, I will merely quote one important result. The maximum 'capacity' of a channel under certain conditions is given by the formula

$$W \log (1 + P/N)$$

where W is the band-width of frequencies employed, N is the average power in this band-width of the noise in the channel, and P the corresponding average power allowed in the signals. The efficiency of actual communication systems, for example those making use of frequency modulation in radio communication, can then be compared with this optimum.

This communication theory has been largely developed by workers at the Bell Telephone Laboratories, especially by C. Shannon.[13] It is quite general and is not confined to electrical methods of communication. A symposium[14] was held in London in 1950 to discuss its numerous ramifications and several statisticians attended who were interested in its important relation with other branches of statistical theory (see also Barnard[11]).

Stochastic Processes in Industry

An older problem in communication engineering associated with the theory of stochastic processes arises, say, in the design of telephone switchboards and is the one of determining 'waiting times' for any

given capacity and density of 'traffic.' But this has so many guises that it is better thought of in the more general terminology of the problem of 'queues,' which the inimitable periodical *Punch* evidently considered (from its review of Mr D. G. Kendall's paper[2]) to be one of the universal problems of our time. Whether the wait is for a disengaged line, or a disengaged shop assistant, or a vacant landing strip at an aerodrome, or a vacant gap in the road traffic, or an available operative to attend to a machine, the wait loses time and money, not to mention our patience. A theoretical and practical study of the stochastic processes involved may help to reduce the amount of time lost.

As an example consider the problem of servicing machines which break down at random times at an average rate r per machine. It is evident that a single group of n operatives to nN machines will be, at least in the absence of any other practical considerations, more efficient than n separate operatives each servicing a separate group of N machines, because the possibility of an operative to one of the latter groups being idle while a machine in another group requires attention is excluded when all operatives are pooled. To illustrate the gain in more detail in a particular case, suppose that the time taken to finish servicing any machine once it has received attention is also random with an average S. Also for simplicity in this example we shall suppose that N is large, but that the average rate of breakdown for all machines remains at a reasonable figure R, say, $(= Nr)$. It might be noticed that our problem is now theoretically equivalent to a queue problem with customers coming in at random to be served, with the operatives representing servers.

The average number of machines (customers) waiting to be served comes out in the case $n = 1$ (one server) as $\alpha^2/(1 - \alpha)$ per server, where $\alpha = RS$, whereas for $n = 2$ it is $2\alpha^3/(1 - \alpha^2)$. Thus the ratio of the average number waiting in the second case to twice the average number for one server is $\alpha/(1 + \alpha)$, indicating the gain in efficiency already referred to. For α, which must be less than 1 if stable conditions are to be maintained, equal to $\frac{1}{2}$ we have $2\alpha^2/(1 - \alpha) = 1$, $2\alpha^3/(1 - \alpha^2) = \frac{1}{3}$, and the gain ratio is as much as 3 : 1. As n further increases the ratio of the expected number of waiting machines to the total expected number in n individual groups with one operative per group tends steadily to zero. This example has, of course, been rather drastically simplified for illustrative purposes, and statisticians interested in this problem should consult a more comprehensive discussion by F. Benson and D. R. Cox.[18]

Apart from the connection already noted of sequential sampling theory with stochastic processes, many sampling problems have been reconsidered in recent years from the stochastic process viewpoint. Thus the problem of sampling from a continuous 'flow' of material has been discussed by G. H. Jowett[20] and the problem of sampling a two-dimensional area (using the idea of a stochastic process over two *spatial* dimensions) by M. H. Quenouille.[5]

It is clearly not only in the *sampling* of a continuous output of some material that a knowledge of stochastic processes may be useful; it will be needed in the statistical analysis and quality control of the material. For example, in the textile industry, whether for cotton, wool, flax, or other fibre, considerable attention has been given to maintaining uniform quality of the yarn. In particular, certain tendencies to periodicity in the thickness of cotton 'slivers' before they are spun into yarn were discussed by G. A. R. Foster at a symposium[24] on time-series held in 1946. It has already been mentioned that a new approach

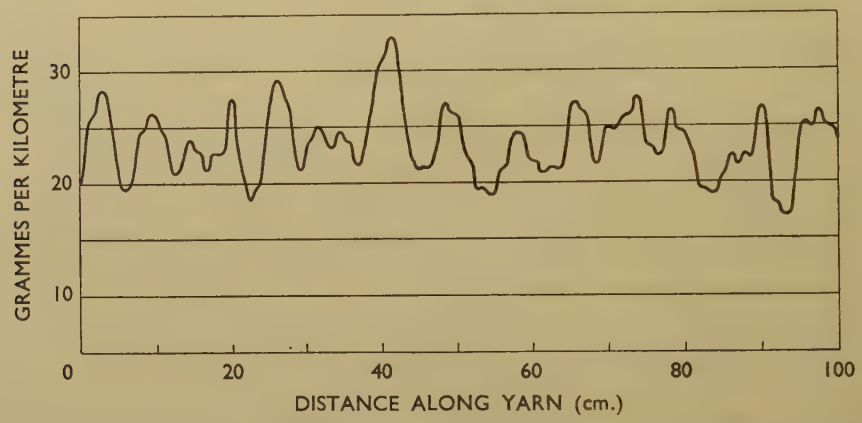


FIG. 2. This chart shows the variation in mass per unit length along a cotton yarn. It was kindly supplied by Mr G. A. R. Foster, who has dealt elsewhere[19,24] with the analysis of such variation. The form of the variation is typical of that shown by a stationary time-series.

to time-series in the last twenty-five years has enabled us to attack such problems statistically with a much greater understanding of what methods to employ.

The Analysis of Economic and Other Time-Series

I have given a technical survey of the statistical analysis of time-series elsewhere[14] and do not want to attempt it here; I do, however, without going into technicalities want to emphasise that the extent to which the statistician can hope to analyse time-series by purely empirical means is now realised to be severely limited. This is because any analysis depends on a theoretical specification or hypothesis of how the data have arisen, and the less detailed the specification available the fewer the conclusions that can be drawn. This is sometimes forgotten in classical statistical problems, where the assumption of a sample of *independent* and *homogeneous* observations is so common that it is often not mentioned explicitly. As an exception, in many experimental situations the lack of independence was not only recognised, but it was effectively eliminated by the randomisation devices introduced by Sir Ronald Fisher. But in time-series, and indeed in stochastic

processes in general, the dependence between the successive observations is usually their most important feature. This has thrown up many new and difficult problems in the theory of statistical inference. However, even when these purely technical problems have been solved, it is necessary to realise that the nature of the dependence has so many possibilities *a priori* that these need first to be drastically restricted in any particular context by theoretical or other sources of information before any analysis is likely to be profitable. Possible exceptions are time-series of the stationary type occurring in some physical or meteorological applications, where the length of series available for study may be more or less unlimited.

In economic and social studies it is rare to have homogeneous series of any length, and any statistical analysis must be closely knit with as full a theoretical specification as possible. Referring especially to this field, Norbert Wiener[16] has made much the same point (p. 35): '. . . the modern apparatus of the theory of small samples, once it goes beyond the determination of its own specially defined parameters and becomes a method for positive statistical inference in new cases, does not inspire one with any confidence, unless it is applied by a statistician by whom the main elements of the dynamics of the situation are either explicitly known or implicitly felt.'

The statistical analysis of economic time-series has thus in recent years been based on rather well-defined hypothetical models of how the variables under study interact with one another. When Udny Yule introduced in 1927 methods of analysis for oscillatory time-series in which the random or stochastic element became incorporated with the future movement of the series, he used the vivid illustration of a swinging pendulum which was being bombarded by boys armed with peashooters. It is not unreasonable to suppose that any random or unpredictable disturbances affecting economic series similarly influence their future movement, and so the theoretical economic models used in the specification automatically become stochastic processes of the kind considered by Yule. One detailed (though somewhat indigestible !) exposition[21] of the statistical methods developed for analysing such types of stochastic series has been published by the Cowles Commission Economic Research Group at Chicago.

Stochastic Processes and the Statistician's Role

The essentially close relation between theory and statistical analysis in the case of economic time-series rather brings to a head the problem of what responsibility devolves on anyone who undertakes statistical analysis. It has always been stressed that the statistician must be fully cognisant of how the figures he is analysing were collected and of any other relevant information, but in the case of stochastic processes it is clear that this ancillary information should also include a very thorough theoretical knowledge of the possible mechanism and structure of the process before any analysis can proceed. We have seen that stochastic

processes may arise in any field of application, and any statistician who finds himself responsible for their analysis must be prepared either to acquire such knowledge himself or to co-operate with someone who has it.

Let us consider an example in the actuarial field, the prediction of population trends. Here it is true that any random or stochastic element affecting separate individuals is practically eliminated if it is the total population size that is of interest. But while the process will appear smooth, its *detailed structure* is a determining factor in its evolution, and it is well known that any empirical extrapolation based on the census figures for the total population is quite inadequate for any but short-term purposes. This is most evident if we consider as an extreme case a fictitious population of young married emigrants who have founded an island colony. The island's birth-rate would at first show large but gradually damped oscillations with a period of about one generation, until the successive generations had had time to merge into each other. Thus more extended extrapolations before the war took full account of (i) the distribution by age and sex of the total population and (ii) fertility and death rates for individual ages. Even this, which it will be noticed takes account of the *instantaneous* detailed structure of the population of individuals, has been recently shown to be insufficient, for it is necessary to recognise the growing custom of planning family size, and hence to try to follow family histories.[22]

Even when random fluctuations are neglected it is worth remembering that they are still there, and that with smaller groups such as some animal populations they may become a crucial factor. Populations are examples of what are called multiplicative stochastic processes, for which fluctuations are cumulative in time. *In appraising the possible size of random fluctuations the statistician must therefore use the theory of fluctuations appropriate to the relevant stochastic process, and this may sometimes be quite different from the 'classical' theory of fluctuations.* Thus for populations with an expected balance of births and deaths relative fluctuations will theoretically tend to be of order $\sqrt{(t/n)}$, in contrast with the classical formula $\sqrt{(1/n)}$, where t is the time in generations and n the mean size of population. For a human population of 50 000 000 and a generation time of, say, 30 years, this is still only of order 1/5 000 after 60 years; but for an animal population of 100 with a generation time of one year it is of order 1/3 after 10 years.

Stochastic Processes in Biology

This theory of population fluctuations is linked with a problem first raised by Francis Galton at the end of the last century in connection with the extinction of family surnames: if each male individual in a population independently has a family containing n sons, where n is a random number following some given distribution, and the sons in turn each have a number of sons following the same distribution, what is the chance of any particular male line becoming extinct? The complete

solution, first obtained by J. F. Steffensen, is a peculiar one. Suppose the probability-generating function of the distribution of n is

$$G(z) = p_0 + p_1 z + p_2 z^2 + \ldots,$$

the probability that $n = r$ being the coefficient p_r of z^r in $G(z)$. Then provided that $p_0 \neq 0$ (otherwise it is obvious that extinction could not occur), the chance of ultimate extinction is the smallest root z_0 of the equation

$$G(z) = z.$$

Moreover, this root z_0 is unity unless the average n is *greater* than unity; even for the United States population in 1920, when the average value of n was 1·145, it was shown by the American actuary Lotka from the statistics of family sizes that the chance of extinction was still nearly 0·9. For example, if for $G(z)$ he substituted the approximate expression $(0{\cdot}482 - 0{\cdot}041z)/(1 - 0{\cdot}559z)$ the above equation became the quadratic equation

$$0{\cdot}482 - 1{\cdot}041z + 0{\cdot}559z^2 = (1 - z)(0{\cdot}482 - 0{\cdot}559z) = 0,$$

giving $z_0 = 0{\cdot}482/0{\cdot}559 = 0{\cdot}86$.

This extinction problem* is important also in the theory of natural selection, since n may alternatively be interpreted as the number of mutant genes in one generation stemming from a mutant gene in the last; as it is unlikely that even a favourable mutation will give an average n much above unity, it will require many occurrences of any such mutation before it is likely to become firmly established in the population.[27]

This last application reminds us that, since the theory of fluctuations in populations is specially important for small populations, it is particularly relevant to all biological population problems involving occasional mutations—the mutant individuals form at first a small population however large the rest of the population may be.

A recent application of the theory of stochastic processes has been to the study of fluctuations in bacterial populations. When the normal bacterial type is placed in an unfavourable environment (for example, a nutrient medium impregnated with bacteriophage) it is possible that a mutant type will arise resistant to this environment. An alternative hypothesis advanced by Sir Cyril Hinshelwood to account for the survival of the bacteria in the new environment is that an actual adaptation of the organism to the environment may lead to survival. Whichever theory may be the correct one (and the latest evidence suggests that neither theory alone is likely to be applicable to all

* For further historical references to this problem see D. G. Kendall's paper 'Stochastic Processes and Population Growth.'[6] The ubiquity of the extinction problem is also indicated (i) by its identification with the gambler's 'ruin' in the historical problems referred to early in this article, and (ii) by its relevance for the epidemiological model treated in the next section.

situations), a study and comparison with observation of the postulated mechanism of growth is evidently necessary. This has been attempted for the mutation theory, and a recent survey is given by Dr P. Armitage.[26]

Stochastic Processes in Medicine

So far I have been taking very much a 'bird's-eye view' of stochastic processes in relation to statistics, and the reader can justifiably complain that the references to applications have been too brief to be other than tantalising. Stochastic processes appear in many branches of medicine also, for example in the study of nervous and cerebral activity or in the study of possible mechanisms of carcinogenesis (see papers by McCulloch & Pitts,[31] and Iverson & Arley[30]). It may, however, be more helpful if I conclude with a single example from the medical field treated in somewhat greater detail. In the discussion following an admirable survey of statistical problems in medicine given at the 1951 Cambridge Conference of the Royal Statistical Society by N. T. J. Bailey[28] I mentioned the probable value of the 'Monte Carlo method' in epidemiological theory. We have seen that it is possible in the growth of very large populations to neglect the stochastic element, but if we do so for epidemics of infectious diseases, especially those whose incidence in the population exhibits a quasi-periodic character, I believe we are in danger of omitting an important factor in their theoretical mechanism.

As an illustration I shall describe the results obtained in a series of fictitious 'measles epidemics,' generated with the aid of random numbers, under conditions simulating a partially isolated group such as a boarding-school. Measles, although not usually a serious complaint, is a favourite infectious disease for study among epidemiologists owing to its relatively simple epidemic character; it confers permanent immunity among almost all those attacked, who are mostly children under the age of 15 (a useful summary of the epidemiology of measles may be found in the late Professor Greenwood's study of epidemics[29]). Notifications are known to exhibit two comparatively stable statistical features. The first is a tendency to biennial periodicity; for example, for Manchester for the years 1917–51 this tendency is quite marked. The second is a seasonal variation; in the case of Manchester for the same years this ranged from 60 per cent. above the average at the beginning of a calendar year to 60 per cent. below in the late summer. It is, however, not at all easy to reconcile these two statistical facts. One may investigate the theoretical consequences of a simple model in which 'susceptibles,' i.e. children who have not yet contracted measles, come in at a given rate and run a risk of infection proportional to the number of infected children present. This model was first shown by Sir William Hamer and later by H. E. Soper[29,33] to be sufficient to produce epidemic waves with a period in time of the right order of magnitude. Unfortunately, in contradiction to what is observed, these

waves damp down until an endemic steady state of infection is reached. If we introduce a seasonal variation in infectivity, a 10 per cent. variation is amply sufficient to produce the observed seasonal variation in notifications; it may be shown that the corresponding theoretical variation in numbers of infected comes out at about 80 per cent. But the seasonal variation forces its own annual period on the waves, and the longer natural period, which corresponds to the observed period, still disappears. Dr Soper believed that the introduction of a definite incubation period of a fortnight counteracted the damping effect but, as first pointed out by E. B. Wilson and J. Worcester[35], he was misled by an inaccurate numerical method. From a study of the transmission of infection in individual households Dr P. Stocks and Miss Mary Karn were led to the hypothesis that some of the children exposed to risk acquire a temporary immunity for about a year without visible contraction of the disease, and suggested that this could contribute to the biennial periodic tendency by protecting unattacked children during the danger period of the following year. But this amended model does not appear on theoretical examination to eliminate the difficulty, and I have begun to suspect that a way out from the dilemma must introduce rather different ideas.

The common feature of the calculations on all these models has up to recently been their straightforward actuarial basis, with no attempt to incorporate the random or stochastic element. I shall refer to such models as 'deterministic' models. But it may prove necessary to recognise the essentially local and hence stochastic nature of infection (random overall variation in infectivity due to weather, etc., will of course also contribute), and on such a new basis continual 'extinction' and replenishment of infected individuals within small groups may create a statistical balance with the damping tendency. In other words, the endemic steady state cannot be attained because it is *stochastically* unstable. To investigate this hypothesis for, say, an entire city is a tall order, for it means studying the stochastic vicissitudes of infections in our theoretical model, which must first be expanded to cover adequately the geographical grouping. It is hoped in due course to carry out such an investigation with the aid of the electronic computer, but in the meantime I have compromised by investigating the similar but simpler problem for a fictitious boarding-school, which is treated as an effectively isolated group of children apart from the influx of children at the beginning of each term.

The precise conditions assumed for the stochastic model are as follows:

(i) *Influx of Susceptibles.*

'Lent term' (1st week)	7
'Summer term' (18th week)	7
'Christmas term' (36th week) ..	23

The numbers are intended to represent a typical case, with much the

greatest influx at the beginning of the school year. It should be noted that immune children are to be ignored.

(ii) *Influx of Infectives.* The entry of infection is assumed to occur by an occasional one or two incoming susceptibles being already infected; the actual number for any term is random, following a Poisson frequency law, for convenience cut off at the value 3. The mean of this distribution is taken proportional to the number of new entrants, but some provisional assumption is also necessary about the seasonal proportion of infectives in the population from which these new entrants are drawn. It seems reasonable, in the absence of a complete stochastic theory for the population outside the school, to base this on the annual oscillations in numbers which follow, as already noted, from an assumed seasonal variation in infectivity on the *deterministic* model. However, it will be seen that the precise assumptions made about the entry of infection are not very crucial provided that some new infection is present from time to time. In the real situation the children would of course not be completely isolated from the rest of the community during term. Moreover, the actual dispersal of the children in the vacations may not only introduce infection through these children but will effectively terminate school epidemics at the end of each term. In the model the vacations are entirely ignored.

(iii) *Infectivity.* The average infection rate per susceptible is assumed to be $\lambda_r = 0{\cdot}01[1 + 0{\cdot}1 \cos(2\pi r/52)]$ per infected person per week, where r is the number of the week in the calendar year, 1st, 2nd, . . ., 52nd. This gives a maximum seasonal infectivity of 10 per cent. above the average at the beginning of the calendar year, and a minimum of 10 per cent. below the average in the middle of the year. The coefficient 0·01 corresponds roughly to the value 1 300 000 originally adopted by Hamer and Soper, based on rates for the increase in numbers of susceptibles and incidence of infection for the whole of London. Its value has been scaled by a factor of 3 000 to be consistent with the very much smaller group represented by a single school, the average number of susceptibles being taken of the order 50 instead of 150 000. This leaves unaltered the approximate period of one and a half years for the gradually damped-out oscillations which follow an initial major epidemic in the deterministic model.

In the stochastic model the chance of infection in a small time-interval δt is taken to be $\lambda_r \mathcal{N}_t S_t \, \delta t$, where $\mathcal{N}_t$ is the number of infectious children at any time t (measured in weeks), and S_t the number of children susceptible to attack.

(iv) *Recovery.* The chance of 'recovery,' by which is meant rather non-infectivity, corresponds to an average infectivity period of a fortnight. For simplicity it is assumed that the chance of recovery per interval δt is $\frac{1}{2}\delta t$, and that an infected child is infectious all the time until such recovery. The actual situation is more complicated; for instance, part at least of the incubation period of about a fortnight is non-infectious and this appears to lead in some local measles epidemics

to a fortnightly periodic structure at the beginning of the epidemic; but we cannot expect any such 'fine-structure' phenomena to appear in our rather crude model. However, it is emphasised that the primary object of the investigation is to examine the self-consistency and possible broad appropriateness of the model; the above oversimplified pattern of infection and recovery actually exaggerates, by a factor of about 2, the damping of the epidemic waves in the deterministic model, but this will not prejudice its relevance if, as in fact is the case, the damping is eliminated in the stochastic model.

(v) *Initial Conditions.* An initial 'epidemic' was started with 100 susceptibles and 5 infectives, these rather arbitrary numbers ensuring, with the high chance of about 0·97, that the artificial series began with a major epidemic.

Complete details of the calculations will not be given, but the way in which they were made can perhaps be indicated. With a process or model developing continuously in time two methods of obtaining artificial realisations of it are possible. With the first, the time is divided up into small intervals and an approximate model proceeding in terms of these small steps is used. This was the method by which the random walk process referred to under the 'Monte Carlo method' could be used to obtain an approximate solution of the diffusion equation, although this equation represents a *continuous* diffusion of particles. This method is not very convenient for the epidemiological problem, for events such as infections multiply rapidly when an epidemic is under way and may be few and far between in the quiescent periods between epidemics. We want in effect our time-scale much finer during an epidemic than at other times. It is therefore more convenient to adopt the second method, in which we determine the random interval between two consecutive events rather than the number of events in a given interval (cf. reference[9]).

Thus the chance of an 'event' in a small interval δt on the above assumptions is

$$a\,\delta t \equiv \lambda_r \mathcal{N}_t S_t\,\delta t + \tfrac{1}{2}\mathcal{N}_t\,\delta t$$

where the first term on the right represents the chance of a new infection and the second the chance that the infectives drop by one. The coefficient λ_r changes only slowly with time and may be treated as temporally constant. $\mathcal{N}_t$ and S_t, apart from the influx at the beginning of terms, cannot by definition change until the next 'event' has occurred, and hence the coefficient a is effectively constant. It follows from the theory developed earlier in this article that the chance that no event occurs in the interval t is the initial probability in a Poisson distribution with mean at, viz. e^{-at}. This defines the so-called 'exponential distribution' for the interval before a further event occurs, and if we write $\mathcal{T} = at$, it will be seen that the distribution of the standardised interval $\mathcal{T}$ on this adjusted time-scale is independent of a. We may thus choose a random interval $\mathcal{T}$ by some convenient method, and then convert

back to the real time scale t by writing $t = T/a$. The method actually used to obtain random intervals was to take $T = \frac{1}{2}(X^2 + Y^2)$, where X and Y are random standardised normal or Gaussian variables available in published tables, but any other method of obtaining a random quantity following the exponential distribution will of course do equally well. At the end of the interval $t = T/a$ we still have to decide *which* event has occurred, but since the relative odds of a new infection and a recovery are $\lambda_r S_t \mathcal{N}_t : \frac{1}{2}\mathcal{N}_t = 2\lambda_r S_t : 1$, we can write $p = (2\lambda_r S_t)/(1 + 2\lambda_r S_t)$ and from reference to a table of ordinary random numbers decide whether the event that has occurred is a new infection. If it is, $\mathcal{N}_t$ goes up by one and S_t down by one; if it is not, $\mathcal{N}_t$ goes down by one and S_t up by one.

A further slight approximation is made at the beginning of each term when, if the last interval goes past this date when conditions change, it is ignored and calculations made afresh from this date with the new conditions. If at any stage $\mathcal{N}_t$ drops to zero, no change can of course occur until a new influx of children occurs.

The start of the calculations is shown in Table I covering the first week, for which $\lambda_r = 0{\cdot}011\,00$. The coefficient λ_r used is that pertaining at the beginning of each week, and it is not changed until the next week is reached. The number of new cases during the week is the decrease in the number of susceptibles, viz. 7.

TABLE I

Random Interval T	a	T/a	p	Random Number	t (weeks)	$\mathcal{N}_t$	S_t
	8·000				0	5	100
1·291 25		0·161	0·688	2952			
	9·534				0·161	6	99
0·077 45		0·008	0·685	4167			
	11·046				0·169	7	98
6·223 40		0·563	0·683	2730			
	12·536				0·732	8	97
0·547 85		0·044	0·681	0560			
	14·004				0·776	9	96
0·210 85		0·015	0·679	2754			
	15·450				0·791	10	95
0·370 00		0·024	0·676	5870			
	16·874				0·815	11	94
0·000 50		0·000	0·674	9268			
	15·340				0·815	10	94
0·276 25		0·018	0·674	2002			
	16·753				0·833	11	93
0·648 65		0·039	0·672	9568			
	15·230				0·872	10	93
2·645 45		0·174	0·672	8243			
					1·046	9	93

By such calculations the mock epidemic series was generated continuously for a total 'time' of 13 years, during which six major outbreaks were obtained with an average period between these epidemics of 125 weeks. The series was evidently 'steady' in the stochastic sense, with

the last epidemics comparable in magnitude with the first. In addition there were four minor outbreaks, in which the entry of infection led to fresh cases but did not lead to a serious epidemic. The results are summarised in Table II, and the course of the largest epidemic (the fifth major epidemic) is shown in Fig. 3. It is stressed that these results are not intended in any sense as a 'fit' to observed data, but statistics for actual measles epidemics in boarding-schools[32] appear to have many similar features if we bear in mind all the further possible complications we have ignored in our model.

TABLE II

Mock epidemic series—summary table

Outbreak	Beginning in Week:	Total Notifications	Percentage Susceptibles Attacked
Major Epidemic			
1	1	98	87
2	105	72	83
3	244	88	85
4	365	67	76
5	504	117	91
6	625	74	87
Minor Outbreak			
1	140	7	18
2	226	7	8
3	556	6	14
4	660	5	15

(On eight other occasions infection entered the 'school' but did not lead to any fresh cases.)

The observed attack rate of presumed susceptibles in all the schools ranged up to 77 per cent. with, however, not such a clear distinction between major and minor outbreaks. (The actual seasonal incidence in the five years covered by the MRC Report[32] seems specially anomalous, with a higher incidence in the summer term than expected from the average in the population at large or in the corresponding mock series. But here it is again stressed that the extent to which the observed statistics can deviate from the average depends on the mechanism producing them,and for only a five-year period, especially as the incidence in different schools is likely to be correlated, considerable fluctuations would be possible.)

To conclude, while it is certainly not claimed that the mechanism of this simple model is sufficient to explain all the observed facts, it is

hoped that this investigation will at least indicate the value of stochastic process models in epidemiological theory. In particular its results, which may be supported by theoretical argument, clearly demonstrate the possibility of a statistical balance between 'extinction' of infectives in a local group and fresh infection from outside. The period will partly depend on the 'isolation' of the local group, but in any case *a major epidemic cannot occur until the concentration of susceptibles has passed*

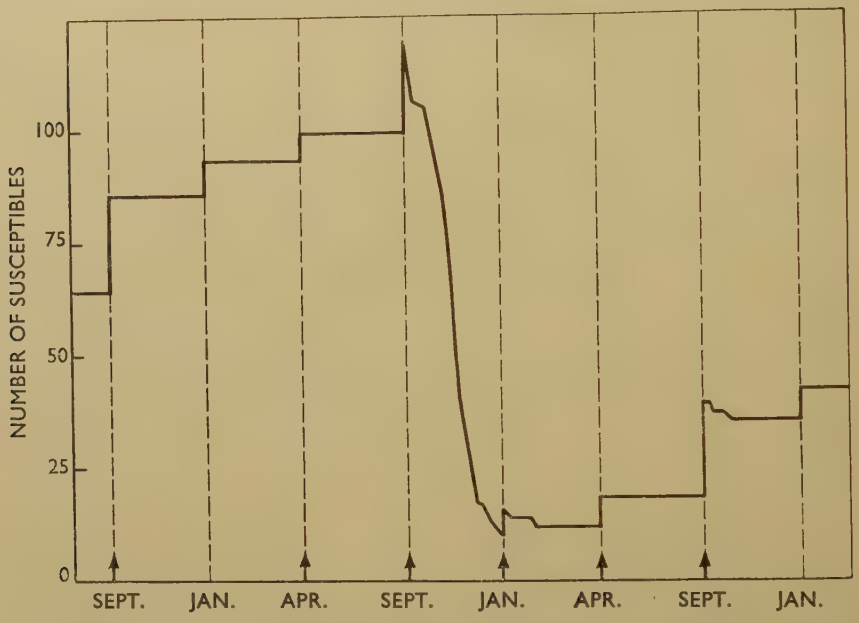


FIG. 3. Extract from mock epidemic series. This graph shows a major epidemic beginning in the autumn of the second 'school' year shown and ending in the New Year. A minor outbreak also occurs the following autumn. The dotted lines indicate the beginning of terms, when there are a number of new entrants to the school; the arrows indicate dates when infection also entered.

its critical threshold value. This critical density of susceptibles is not altogether a new notion to epidemiologists but it is suggested that only with the stochastic approach, in which a smaller density ensures local extinction, can it acquire proper theoretical justification.

I am greatly indebted to Mrs A. Linnert for carrying out the detailed computation for the artificial epidemic series.

REFERENCES AND BIBLIOGRAPHY

This short list of papers and books is confined mainly to applications referred to in the text. For the convenience of readers it has been classified according to the field of application, though this necessitates some cross-referencing. Many further references will of course be found in these publications.

I. *General*

[1] Arbous, A. G., & Kerrich, J. E. (1951). 'Accident Statistics and the Concept of Accident Proneness,' *Biometrics*, **7**, 340.

[2] Kendall, D. G. (1951). 'Some Problems in the Theory of Queues,' *J. Roy. Statist. Soc.* B, **13**, 151.

[3] Lundberg, O. (1940). *On Random Processes and their Application to Sickness and Accident Statistics*. Almqvist & Wiksells, Uppsala.

[4] McKendrick, A. G. (1914). 'Studies on the Theory of Continuous Probabilities, with Special Reference to its Bearing on Natural Phenomena of a Progressive Nature,' *Proc. Lond. Math. Soc.* (2), **13**, 401.

[5] Quenouille, M. H. (1949). 'Problems in Plane Sampling,' *Ann. Math. Statist.*, **20**, 355.

[6] Symposium on Stochastic Processes (1949). *J. Roy. Statist. Soc.* B, **11**, 150.

[7] Wald, A. (1947). *Sequential Analysis*. John Wiley, New York.

II. *Monte Carlo Method*

[8] Hammersley, J. M. (1950). 'Electronic Computers and the Analysis of Stochastic Processes,' *Math. Tables and Other Aids to Comp.*, **4**, No. 29, 56.

[9] Kendall, D. G. (1950). 'An Artificial Realisation of a Simple "Birth-and-Death" Process,' *J. Roy. Statist. Soc.* B, **12**, 116.

[10] National Bureau of Standards (1951). *Monte Carlo Method*, Applied Mathematics Series, **12**. Washington, DC.

III. *Physics and Communication Engineering*

[11] Barnard, G. A. (1951). 'The Theory of Information,' *J. Roy. Statist. Soc.* B, **13**, 66.

[12] Chandrasekhar, S. (1943). 'Stochastic Problems in Physics and Astronomy,' *Reviews of Modern Physics*, **15**, 1.

[13] Shannon, C. E., & Weaver, W. (1949). *A Mathematical Theory of Communication*, University of Illinois Press, Urbana.

[14] Symposium on Information Theory (1950). Ministry of Supply, London.

[15] Wiener, N. (1949). *The Extrapolation, Interpolation and Smoothing of Stationary Time-series, with Engineering Applications*. John Wiley, New York.

[16] *Idem.* (1948). *Cybernetics*. John Wiley, New York.

IV. *Industry* (See also I[1,2,3,5,7])

[17] Barnard, G. A. (1946). 'Sequential Tests in Industrial Statistics,' *J. Roy. Statist. Soc.* B, **8**, 1.

[18] Benson, F., & Cox, D. R. (1951). 'The Productivity of Machines requiring Attention at Random Intervals,' *J. Roy. Statist. Soc.* B, **13**, 65.

[19] Foster, G. A. R. (1951). 'Fibre Motion in Roller Drafting,' *Shirley Inst. Mem.*, **25**, 51, and *J. Text. Inst.*, **42**, T335.

[20] Jowett, G. H. (1952). 'The Accuracy of Systematic Sampling from Conveyor Belts,' *Applied Statistics*, **1**, 50.

V. *Analysis of Economic and Other Time-Series* (See also III[14,15,16] and IV[19])

[21] Koopmans, T. C. (editor) (1950). *Statistical Inference in Dynamic Economic Models*. John Wiley, New York.

[22] *Report of the Royal Commission on Population* (1949). HMSO, London.

[23] Slutsky, E. (1927). 'The Summation of Random Causes as the Source of Cyclic Processes,' *Problems of Economic Conditions*, ed. by the Conjuncture Institute, Moscow, **3**, No. 1 (later reprinted in *Econometrica*, **5** (1937), 105).

[24] Symposium on Autocorrelation in Time-Series (1946). *J. Roy. Statist. Soc.* B, **8**, 27.

[25] Yule, G. U. (1927). 'On a Method of Investigating Periodicities in Disturbed Series, with special reference to Wolfer's Sun-spot Numbers,' *Philos. Trans.* A, **226**, 267.

VI. *Biology* (See also I[4,5,6] and VII)

[26] Armitage, P. (1952). 'The Statistical Theory of Bacterial Populations subject to Mutation,' *J. Roy. Statist. Soc.* B, **14**, 1.

[27] Fisher, R. A. (1930). *The Genetical Theory of Natural Selection.* Oxford University Press.

VII. *Medicine*

[28] Bailey, N. T. J. (1952). 'The Scope of Medical Statistics,' *Applied Statistics*, **1**, 149.

[29] Greenwood, Major (1935). *Epidemics and Crowd Diseases.* Williams & Norgate, London.

[30] Iversen, S., & Arley, N. (1950). 'On the Mechanism of Experimental Carcinogenesis,' *Act. Path. and Microbiol. Scand.*, **27**, Fasc. 5, 773.

[31] McCulloch, W. S., & Pitts, W. (1948). 'The Statistical Organisation of Nervous Activity,' *Biometrics*, **4**, 91.

[32] Medical Research Council (1950). *Epidemics in Schools*, Special Report No. 271, HMSO, London.

[33] Soper, H. E. (1929). 'The Interpretation of Periodicity in Disease Prevalence,' *J. Roy. Statist. Soc.*, **92**, 34.

[34] Welch, J. D., & Bailey, N. T. J. (1952). 'Appointment Systems in Hospital Outpatient Departments,' *Lancet*, **1**, 1105.

[35] Wilson, E. B., & Worcester, J. (1945). 'Damping of Epidemic Waves' and 'The Spread of an Epidemic,' *Proc. Nat. Acad. Sci. Wash.*, **31**, 294 and 327.

QUESTIONS AND ANSWERS

Readers are invited to submit questions.

Question

My chairman's office has its own telephone switchboard of five lines, of which two are reserved for the exclusive use of the chairman for outgoing calls only. The remaining lines are used for all incoming calls, and for outgoing calls by the chairman's personal assistants and so on. This, I feel, results frequently in incoming calls being held or rung as 'number engaged' and in delays in making outgoing calls, even when the chairman's lines are not actually in use. Surely there must be a way of deciding how many lines, if any, should be reserved in order to get full use out of the switchboard and yet to ensure that the chairman will nearly always find a line free for his outgoing calls.

Answer

The problem of assessing the risk that all lines will be occupied when a call is made and other problems like it have been studied, notably by A. K. Erlang. The fundamental formula gives the probability P_n that a call will find all of n lines occupied as

$$P_n = \frac{y^n/n!}{1 + y + y^2/2! + y^3/3! \ldots y^n/n!}$$

where y is the traffic density (the number of calls per average duration of a call). The formula assumes that a statistically steady state has been established, but if a steady state cannot be assumed, the use of the y for a busy period will always result in an over-, and therefore safe, estimate of P_n. The present problem might be tackled in the following way.

1. Establish the traffic densities y. If the chairman's lines are sometimes to be reserved for outgoing calls only, the Post Office may require them always to be so reserved; it will then be necessary to assess traffic densities for incoming and outgoing calls separately.

2. For any given period of the day calculate the risk P_3 that 'office' calls (with densities of y_i inward and y_o outward) are blocked when three lines only are used. Next calculate the risk, with the chairman's outward calls (y_o') added, when five lines are used. This enables the gain in efficiency to the office to be estimated if both the extra lines are made available for both inward and outward calls. In the five-line system the chairman's risk of being blocked is the same as for anyone else, but given priority, his time of waiting is less.

3. If, alternatively, the two extra lines are to be used for outward calls only, then the blocked outward calls on the three-line system plus the chairman's outward calls (that is, a density of $y_o P_3 + y_o'$) are fed to the two-line system, resulting in a new blocking risk P_2. This compares with P_3 above for assessing the gain to the office. The risk to the chairman changes from zero to P_2.

4. If one of the outward lines is still reserved for the chairman's outward calls a similar calculation can be made, remembering that the addition to $y_o P_3$ for traffic density on the remaining outward line is only those of the chairman's outward calls which require two lines at once.

Further information may be obtained from 'The Life and Works of A. K. Erlang,' by E. Brockmeyer, H. L. Halstrom, and A. Lensen, *Transactions of the Danish Academy of Technical Science*, 1948, and *Probability and its Engineering Uses* by T. C. Fry, Van Nostrand, New York. E. D. VAN REST

Question

I have recently read in the *Journal of the American Statistical Association*, June 1952, p. 281, an article by Neva R. Deardorff, entitled 'Prepaid Medical Care as a Source of Morbidity Data.' Where can I obtain particulars about morbidity statistics in Great Britain and America having reference to *all* medical services at physicians' surgeries as well as on their rounds (no hospital statistics), with specification of different diagnoses? What is the situation in any other country?

Answer

No comprehensive morbidity statistics exist for England and Wales. Until recently sample enquiries have been made into the incidence of illness, incapacity, and medical consultations in England and Wales, and the results, including some by diagnoses, are published in the quarterly report of the Registrar General. New claims to sickness benefit and injury under National Insurance Acts are given for each week in the weekly report of the Registrar General. This report also includes the numbers of cases of certain infectious diseases reported in different areas and incidence rates are shown in the annual report of the Registrar General. Data on the incidence of tuberculosis and venereal diseases are given in the annual report of the Chief Medical Officer of the Ministry of Health.

No material is available that would give an adequate estimate of the services rendered by the average general practitioner. Some idea of what these consist in may be obtained from 'A Year of General Practice: A Study in Morbidity' by John Fry, *British Medical Journal* 2nd August 1952, p. 249; and 'A Year of General Practice' by C. A. H. Watts, *Supplement of the British Medical Journal* 30th August 1952, p. 115.

I know of no comprehensive official morbidity statistics for America and our knowledge depends on enquiries such as the one quoted. I do not know the situation for other countries. W. J. MARTIN

NOTES AND COMMENTS

THE British Standards Institution has long been recognised as an important centre for promoting the application of statistics in industry. To be really satisfactory and effective, the specifications entering into standards, especially standards containing sampling clauses, must take statistical considerations fully into account, and a well-conceived standard can have an influence in encouraging quality control in industry. Moreover, the BSI has a prestige in industry, so that any publications it issues in its field are highly regarded. Before the war the BSI established a Committee on Statistical Methods in Standardisation and Specification which had a considerable influence in the development of what was then a new movement. The chief visible achievements of this committee are some BSI publications which it sponsored, notably Professor E. S. Pearson's *The Application of Statistical Methods to Industrial Standardisation and Quality Control* and Dr B. P. Dudding's and Mr W. J. Jennett's *Quality Control Charts*. After the war this committee, for a variety of reasons, became moribund; but people both inside and outside the Institution recognised that the BSI had by no means finished its work for statistics in industry. Accordingly, under its new organisation the Institution has appointed a new Technical Committee to take up and carry on the work. Some British Standards show admirably what can be done to make a specification statistically sound, many others show how much remains to be done; and it is by stimulating and advising other committees responsible for standards that this new committee can have most influence in the near future. Fundamentally, however, such influence can only be exerted through giving detailed and expert services such as a large committee cannot give, and it is an important first step that the Committee has been able to appoint a small panel of gentlemen who are competent and in a position to give what is required. The public-spirited members of this panel are Mr R. F. H. Banister, Dr B. P. Dudding, Dr H. O. Hartley, and Mr A. J. H. Morrell.

★ ★ ★

There will be a wide welcome for *The Journal of Industrial Economics*, which is described as an 'Anglo-American Journal especially devoted to the economic problems of industry and commerce' and 'offers a medium for the academic industrial research work increasingly being carried on both in American and Britain, and also for the material arising inside business itself—important internal memoranda on problems facing a particular business or industry, which are at present lost to a wider audience.' The first number appeared in November 1952 and contains, among others, an article by Mr Frank A. Friday on 'The Problem of Business Forecasting.' The editor is Mr P. W. S. Andrews of Nuffield College, Oxford, and the publishers are Basil Blackwell, Oxford. The annual subscription is one guinea or $3 post free for three issues, and the price of a single issue is 10s. or $1.40 post free.

INDUSTRIAL APPLICATIONS SECTION OF THE ROYAL STATISTICAL SOCIETY

List of Meetings

BIRMINGHAM GROUP

24th September 1952. Designed Experiments and their Relationship with Statistical Quality Control: M. Milbourn
29th October 1952. Statistical Quality Control in a Laboratory: B. J. A. Martin
26th November 1952. Analysis and Cost of Labour Turnover: J. Murdoch

GLASGOW GROUP

15th October 1952 (Inaugural meeting). Control Chart Techniques as an Aid to Supervisory and Executive Staff: B. P. Dudding
19th November 1952. Quality Control Applications: W. A. Bennett

LEICESTER GROUP

9th October 1952. Some Applications of Analyses of Variance to Industrial Problems: A. Hardy and A. Bramley
13th November 1952. Patrol Inspection—Economic and Statistical Considerations: M. J. Moroney
4th December 1952. Statistical Methods in the Hosiery Industry: A. Glaskin

LONDON GROUP

3rd October 1952. The Design and Analysis of 2^n Factorial Experiments with Fractional Replication: G. W. Sears
7th November 1952. Discussion of Hand Operated Punched Card Systems: Miss J. Keen and L. T. Wilkins
5th December 1952. The Statistical Approach to Congestion Problems in Industry: J. Murdoch

MERSEYSIDE GROUP

16th October 1952. The Scope of Industrial Statistics: W. J. Jennett
18th November 1952. Transformations used in Statistical Analysis: R. L. Plackett
16th December 1952. Tolerances: G. A. Barnard

NORTH-EASTERN GROUP

15th October 1952. Quality Control in the USA: E. D. van Rest
19th November 1952. Accident Records: H. Campbell
17th December 1952. Economic Statistics: E. D. McCallum

SHEFFIELD GROUP

25th September 1952. Graphs and Correlations: D. G. Beech
30th October 1952. The Statistically Based Production Control of Heterogeneous Output: S. Beer
27th November 1952. Some Problems on the Sampling of Airborne Dust in Coal Mines: B. A. Maguire

SOUTH WALES GROUP

2nd October 1952. Statistics Research: A. W. Swan
6th November 1952. Statistical Techniques to determine Optimum Conditions: G. E. P. Box
11th December 1952. British Earnings 1938–50: E. J. Cleary

★ ★ ★

The inaugural meeting of the Glasgow Group, listed above, was held in the Institution of Engineers and Shipbuilders in Scotland. Mr. W. A. Bennett was elected Chairman of the Group, Mr. C. E. K. Scoular Vice-Chairman, and Mr. Andrew Kirk Secretary and Treasurer. A programme of six meetings was arranged for the session; so far these have been well attended.

BOOK REVIEWS

Oscar Kempthorne: *The Design and Analysis of Experiments.* New York: John Wiley and Sons; London: Chapman and Hall, 1952. Pp. xix + 631. 68s.

This is a good book; it deals comprehensively with experimental designs used in agricultural research and introduced and developed largely by the Rothamsted school, e.g. randomised blocks, Latin squares, etc., various types of incomplete blocks with one or more restrictions, and factorial designs. The book also deals with confounding and fractional replication, and includes much of the author's own work on fractional replication and lattice designs. The main emphasis is on the basic statistical concepts, the mathematical theory, and the general structure of experimental design. Mathematical justification is given at all stages, and there are a treatment of the method of least squares, a treatment of general linear hypotheses, and a discussion on such matters as the principles of randomisation and the validity of the methods of analysis.

Although there are basic differences between agricultural and industrial research work the statistical techniques used in agriculture are, to a large extent, applicable also to industrial work, and in this respect the book will be of value to the industrial statistician.

The author has three types of readers in mind: 'the experimenter who needs to understand the theory and practice of statistical design; the consulting statistician who wants a description of the basic theory and methodology of the design; and the mathematical statistician who is concerned with the validity of the application of the statistical techniques to the analysis of data.' All such readers will certainly find the book of value, but it is definitely not a book for the beginner, as a fair knowledge of mathematics and statistical analysis is required to obtain a reasonable understanding of the book.

References for further study are quoted at the end of each chapter, and the full list of contents and the index add to the usefulness of the book.

O. L. DAVIES

A. Lester Boddington (10th Edition by A. R. Ilersic): *Statistics and their Application to Commerce.* London: Pitman and Sons, 1952. Pp. xiv + 451. 25s.

This text-book has been completely rewritten for its tenth edition. It is primarily for those studying at home to take statistics as a subsidiary subject in a professional examination, though a successful attempt has been made to prevent the treatment of methods being purely mechanical. In addition to sections on the usual methods of describing large masses of figures, methods ranging from diagrams and averages to regression and index numbers, there are sections on quality control, sampling, and tests of significance. There are chapters on sources, current index numbers, and on some survey organisations.

The chapters on the measurement of averages and variation are clear, including worked examples showing the use of arbitrary origins in calculating

the arithmetic mean and standard deviation, and showing that the calculation of the two together is a natural process, not to be divided into separate compartments called respectively 'the calculation of the arithmetic mean' and 'the calculation of the standard deviation.' The treatment of index numbers, too, is clear; and new students are not likely to confuse 'weight' with avoirdupois when the primary discussion concerns price relatives. The 'weight' is referred to as the importance of the article, and there is a supplementary statement, in brackets, that six times as important means that six time as much is spent on it.

I am not so satisfied with the treatment of regression and correlation, which seems more complex than is necessary for an elementary book yet must leave the student uncertain of how to tackle the typical examples of commerce and industry. In the example given, involving thirty items, working from the true means is simple because those true means happen to be simple numbers (104 and 2·5) with 16 and 1·5 as the biggest deviations therefrom. What is the student expected to do if he has information about 300 items and if the averages are not simple numbers? The only other way shown is that of measuring from the natural origins, and the answer depends upon small differences between large numbers so that minor inaccuracies of working can easily produce misleading, or even absurd, results. Should not the student be shown that a convenient way of estimating the correlation coefficient in typical cases is the preparation of a two-way frequency table and the use of arbitrary origins?

I am not certain that a standard text-book on statistical methods should also include a section on sources. The latter 'date' so much more quickly than the methods that there seems scope for a separate, cheap volume on sources which can conveniently go into fully revised editions more frequently than the book on methods.

This is, however, a book to be recommended on the whole to all new students of statistical methods, whether they are studying at home or not.

H. S. BOOKER

M. G. Kendall (Editor): *The Sources and Nature of the Statistics of the United Kingdom.* Volume I. Published for the Royal Statistical Society. London and Edinburgh: Oliver and Boyd, 1952. Pp. viii + 352. 21s.

The flow of British economic statistics has increased so much in recent years that it threatens to become a flood. Even the specialist now finds it difficult to keep up to date with all the figures that are coming out, and the changes in definition and coverage that seem inevitably to occur at frequent intervals make the task of absorption and interpretation especially difficult. Pity the non-specialist user in commerce and administration, and the overburdened student for whom the study of economic statistics is merely a subsidiary and perhaps optional subject! How grateful they will be to the Royal Statistical Society for commissioning experts to write short essays explaining to us in summary what is available, where it can be found, and how to avoid misinterpretation. And gratitude to the Society will be even greater for collecting these essays, originally published in its *Journal*, all together in one volume. This first volume (the task is not yet complete—further essays are being published) covers articles on the Censuses of Production and Distribution, Oversea Trade, Agriculture, Labour Statistics, Coal Mining Industry, Rubber, Cotton Industry, Brewing Industry, Oils and

Fats, Iron and Steel Industry, Tobacco, Electricity and Gas, Sugar Industry, Motor Industry, Chemical Industry, Rail and Road, Housing Statistics, Co-operative Trading, and Publishing.

In an enterprise of this kind considerable discretion must necessarily be left to each author, and the scope, method of treatment, and standard of the articles, vary a good deal. Consequently the value of the final product also varies. For example, agriculture gets forty pages, with a complete and detailed description of sources, a discussion of the most important figures themselves (giving an excellent statistical picture of British agriculture), and an extensive bibliography of 169 items. By contrast, labour statistics are dealt with in twelve pages, discussion is confined almost completely to Ministry of Labour figures, there are very few actual statistics, and the bibliography comprises only nine items.

One must be careful not to treat these articles as the last word on the subjects with which they deal. The flow of new statistics is so great that anything written is almost certain to be out of date before it is published. For example, the excellent article on housing statistics by Miss Bowley necessarily takes no account of the recently published sample results of the Census of Population for 1951. Dare one hope that the Society will be able to persuade the authors to bring their articles up to date when later editions of this volume are published ?

This is obviously not a book to be read straight through. But it will be invaluable for reference to those who use British economic statistics. How much easier such reference would be if the Society had managed to persuade someone to compile an index ! However, these are minor criticisms. We shall all look forward to Volume II.

E. DEVONS

L. H. C. Tippett: *The Methods of Statistics.* London: Williams and Norgate; New York: John Wiley and Sons, 4th Edition, 1952. Pp. 395. 38s.

This is a revised edition of the book written in 1931, especially for experimenters, 'to expound Fisher's methods in comparatively simple terms.' Since the book first appeared there has been a growth of the subject and some shift in the ways of approach to many of the problems. The book had twice previously been revised to keep position in this growth. In the present edition the rewriting of several of the chapters, the division of others, and the addition of a chart to correspond with each statistical table, account for over one hundred new pages. It is no criticism to say that the result is a mixture of old and new, for though there is certainly room for a new presentation of the subject with a more unified method of approach such a book does not exist, and the need meanwhile is for a book such as the present one where the reader may find a straightforward explanation of why and how he should use a particular test or form of analysis under the name by which he already knows it.

The new pages deal among other things with time series, ranking methods, sampling, fitting orthogonal polynomials, transformations of variates, and discriminant functions. It is disappointing to find so little under the heading Principles of Experimental Arrangement; pertinent material appears, however, in other sections and the bibliography is extensive. Probit analysis deserves more than the passing mention it gets. The book is well produced in clear type on paper that gives no glare in artificial light; the only shock comes from the price.

E. D. VAN REST

EXPERIENCES OF CORRELATION ANALYSIS

DONALD G. BEECH

The British Ceramic Research Association

Dr Beech uses some experiences in the field of ceramics as pegs on which to hang discussions of practical points that arise in correlation analysis. He deals with the dangers of extrapolation, the use of correlation analysis to test the validity of a theoretical relationship, and the estimation of a constant of proportionality from data.

The graphical presentation of numerical results is such a common feature of scientific papers that there is a danger that its full significance may sometimes be forgotten. It is well known that the graph is usually a much more effective method of demonstrating results than a table, yet the true function of a graph, as distinct from a chart, lies somewhat deeper than this. Underlying all graphs there is the conception of a correlation between the variables that have been plotted. Hence, whenever a graph is drawn there is the possibility of correlation analysis. It is often unnecessary to perform this analysis since certain results 'speak for themselves,' but there are many occasions when a full analysis, though laborious, is very rewarding.

There are two main types of correlation and both are encountered in scientific work. In the first type there is reason to believe that the two measured quantities are connected by a strictly definable functional relationship, and that the points on the graph depart from this relationship only by reason of experimental error in one or both measured quantities. The mathematical treatment is much simplified if the relationship is believed to be linear. Then, according to the distribution of the errors, one or other of the regression lines may be calculated for the purpose of estimating one variable from the measured value of the other.[1] The linear functional relationship between the true values of the variables may be calculated if there is sufficient information about the experimental errors.[2] An example of two quantities thus functionally related would be the pressure and temperature of a constant mass of gas at constant volume.

In the second type of correlation the variables, though correlated, are not functionally related. Experimental error may be present, but it is not the sole cause of departure from a clear-cut relationship. Again the treatment is simpler if the plotted points are distributed roughly round a straight line. In such cases the appropriate regression line is used to estimate one variable from the measured value of the other one, and no meaning can be attached to a single line of best fit. The correlation coefficient is important here since it indicates the strength

of the relationship between the two variables. An example of such a correlation is that between the Rockwell hardness and tensile strength of metals.

Having indicated briefly the place of correlation analysis in scientific work, this article will now deal with some of the problems encountered in practice. It makes no claim to be comprehensive but merely illustrates some of the principles and some of the pitfalls.

Extrapolation

Graphs are often used to enable values of one variable to be estimated from observed values of the other. When the value required is within the range of values from which the graph was deduced this procedure is known as interpolation. Provided that the assumption of correlation between the variables is justified, interpolation is logically sound. When the value required is outside the above range the procedure is termed extrapolation, and is open to severe criticism. It may be argued that extrapolation is never justifiable, because it entails making predictions in a field that has not been explored and often relies too much on the extreme points on the graph, such points being frequently of low accuracy. On the other hand, there are cases where there are theoretical reasons for believing that the relationship which has been verified over the range covered by the graph also holds over the range in which it is proposed to extrapolate. In such cases it seems reasonable to allow extrapolation, especially if the range of extrapolation is narrow. The greater the range of extrapolation the less is the accuracy.

Frequently it is difficult to find a valid reason why a relationship should be extrapolated, but there appears to be no reason why it should not be. In such circumstances it is foolish to extrapolate when the desired information could have been obtained by direct measurement without undue trouble. In practice extrapolation may sometimes save so much labour and time that the risk of a false prediction is taken.

Extrapolation is really indefensible when careful consideration would reveal reasons why the relationship between the variables should show a discontinuity in the range of extrapolation. An example may be quoted. The investigation was concerned with the process of casting a clay body in a plaster mould. It was desired to determine the moisture content of the cast immediately after the clay suspension had been poured from the mould. It was impossible to remove the cast from the mould immediately after pouring without damaging it, and it was desired to keep the cast intact. It could, however, be removed without damage after one hour and weighed. The procedure adopted was to remove the cast after one hour, weigh, replace in the mould, remove and weigh again after a second hour, etc., so getting the weights at successive intervals after pouring off the clay slip. Finally the cast was dried and weighed again, and the moisture content at each of the four times was calculated. The moisture contents were plotted against the time of standing; a typical graph is shown in Fig. 1, where the different

curves refer to different durations of casting. It may be seen that increased time of standing results in lower moisture content, chiefly owing to the loss of water to the porous mould. It had been assumed that, because of the small curvature of the line joining the four points, it could be extrapolated to zero time to give the moisture content immediately after pouring. Unfortunately, during the first hour there is a

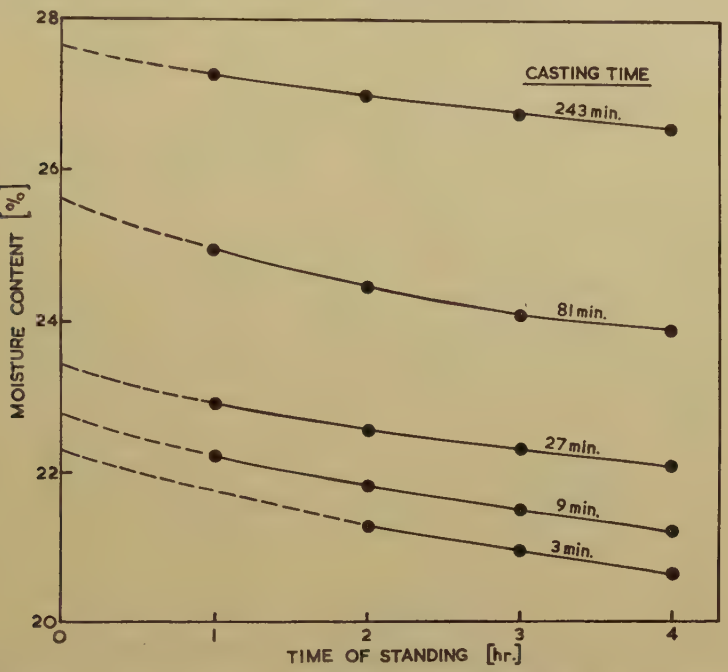


FIG. 1. Moisture content of cast article measured at various times after the end of the casting process.

marked change in the conditions. At first the cast cannot easily be removed; the reason is that it is then in intimate contact with the mould and sticks to it. At the end of the first hour it can easily be removed, so it is obvious that intimate contact must have been lost. Now the rate of transfer of water to the mould would be expected to be far greater during the time of intimate contact than at any time after that contact becomes less close. Hence one would expect a fairly sharp discontinuity during the first hour; and the extrapolation is not valid. In point of fact, an independent experiment showed that the amount of water transferred to the mould during that first hour would account for a far greater change in moisture content than that predicted by extrapolation.

Testing a Theoretical Relationship

An interesting use of correlation analysis is in the testing of theoretically deduced relationships. The example chosen is remarkable in showing a very high degree of correlation; it also shows some pitfalls.

It concerns the relationship between the specific gravity and the water content of a clay suspension. For convenience 20 times the specific gravity was plotted as x to facilitate subsequent use of the relationship. The water content W is the weight of water divided by the weight of clay. If there were no condensation between the clay particles and water the volume relations would be expected to be additive. In such circumstances the relationship between W and x may be shown to be nearly hyperbolic, and to facilitate subsequent testing it may be written in the form

$$W(x - 20) = 20 - x/S$$

or $y = 20 - x/S$, where $y = W(x - 20)$ and S is the specific gravity of the dry clay. If there were condensation the last term would probably be of the form x/Z, where Z is not equal to S and is possibly not even a constant.

The validity of the simple equation was tested by plotting y against x; only six sets of experimental results were determined, but every effort was made to get as much accuracy as possible with the normal technique. (The determination of such a small number of results is not

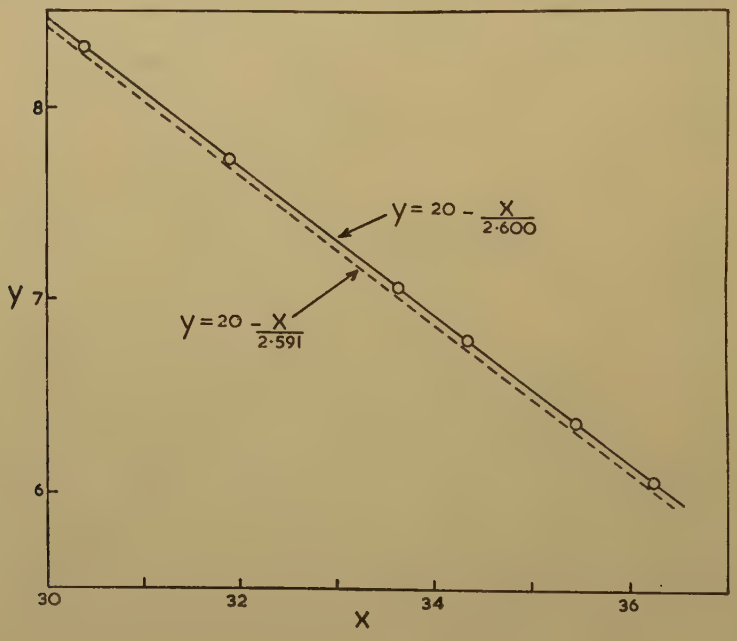


Fig. 2. Relationship between physical properties of clay-water suspension.

defended; the analysis had to make the best use of the data available.) The results are given in Table I and Fig. 2. The graph is clearly linear; are the points in agreement with the equation theoretically deduced?

Since y contains both variables it was considered to have more error than x. The regression line of y against x was therefore calculated, thus assuming all the errors to be in y and using the principle of least squares to minimise these. (Since the correlation coefficient was

TABLE I

Data for correlation analysis on properties of clay suspension

x	y
30·37	8·317
31·90	7·735
33·64	7·066
34·35	6·788
35·45	6·365
36·24	6·058

0·999 993 it would have made no appreciable difference if the other regression line had been taken.) The calculated regression line was

$$y = 20{\cdot}015 - x/2{\cdot}597$$

The procedure first used to test the relationship was as follows. The intercept on the y axis is 20·015 with a standard error of 0·024; consequently this intercept does not differ significantly from the theoretical value 20. The reciprocal gradient, 2·597, is an estimate of the specific gravity of the clay if the theoretical equation applies. This estimate has a standard error of 0·005. The specific gravity was estimated from 13 determinations by the pycnometer method as 2·591 with a standard error of 0·0014. The two estimates do not differ significantly and hence the theoretical gradient is confirmed. It was therefore concluded that the regression line does not differ significantly from the theoretical line

$$y = 20 - x/2{\cdot}591$$

This equation gives the broken line in Fig. 2, and it may be seen to run roughly parallel to the points but not through them. It does not seem to be a good fit; the difference between the points and line looks sufficiently regular to be significant. The reason for the apparent paradox is clear when the meaning of the significance tests is more closely examined.

The regression line may be written in general

$$y = a + b(x - \bar{x})$$

where a and b are calculated from the experimental data and $\bar{x}$ is the mean value of x. The theoretical line is

$$Y = A + Bx$$

where A and B are known from other evidence. The test of the gradient compares b and B. The test of the intercept compares $(a - b\bar{x})$ and A. The test of the gradient is satisfactory. The test of the intercept may here be regarded as a comparison of the positions of the two lines at the value $x = 0$. Now the variance of a point on the regression line is given by

$$V(y) = v\{1/n + (X - \bar{x})^2/\Sigma(x - \bar{x})^2\}$$

for the point $x = X$, where v is the residual variance about the line and n is the number of experimental points. The first term corresponds to the error in a and the second term to the error in b. The intercept at $X = 0$ gives in the present instance a high value of $(X - \bar{x})^2$ and consequently a high value of $V(y)$. (This demonstrates the high standard error, and hence low accuracy, of a point predicted by wide extrapolation.) Hence the comparison of positions is not very sensitive at this point.

The lowest standard error of a point on the regression line is obtained when $X = \bar{x}$. Therefore the most sensitive test of the positions of the two lines is at this point, this test being equivalent to a comparison of a and $(A + B\bar{x})$. This procedure could be supplemented by comparison of the positions at other values of x within the experimental range.

A similar conclusion may be reached in another way. The test of the gradients shows whether a line through $(\bar{x},\bar{y})$ with gradient B is consistent with the observed points; the test of the intercept shows whether the line through the point $(0,A)$, with another value of the gradient determined by the points themselves, is also consistent with the data. When both these tests fail to show significant differences, as in the present case, it is not always safe to infer that the line $Y = A + Bx$ would be satisfactory. The only safe plan is to test the closeness of the fit of $Y = A + Bx$ at a number of points in the experimental range; in some circumstances a single analysis of variance may be employed with advantage.

This was accordingly done. The difference in the ordinates of the regression line and the theoretical line at the point $\bar{x}$ was compared with the standard errors of the two lines at this point. A highly significant difference was found and this was confirmed by tests of the difference between the lines at other values of x. (If the theoretical line had been known with absolute precision these tests could have been combined in a single analysis of variance.) Hence these results support the subjective judgment that the theoretical line is not a good fit.

Although the simple theory must be rejected, the experimental results are consistent with either the theoretical intercept or the theoretical gradient, but not both. Now there were grounds for believing that condensation might affect the gradient but not the intercept. The best line through the point (0,20) was therefore calculated and found to be

$$y = 20 - x/2{\cdot}600$$

This equation has already been shown by the intercept test to be consistent with the points, and its graph (the full line in Fig. 2) gives visible support to this finding. The value of the gradient may be explained by a small amount of condensation.

This example has been quoted in detail because it shows the danger of applying simple tests without fully considering their meaning, the danger of relying on the accuracy of wide extrapolation, and the usefulness of plotting a calculated line so as to check the adequacy of the fit.

Estimation of Factor of Proportionality

The use of graphs for estimating the values of one variable from the measured values of another, simple though it is as a rough method, becomes complex when we seek to refine it. For a complete solution of the problem one may need precise estimates of the error to which each measurement is subject, although in certain cases an approximate estimate of the error distribution is sufficient. Here only one special case of estimation is dealt with, namely the estimation of y from x, or x from y, when x is virtually free from error and there are adequate theoretical and practical grounds for expecting the relationship between x and y to be a straight line passing through the origin. The problem is to determine the value of the proportionality factor from a number of measured pairs of values of x and y.

This factor is often determined by plotting the values on a graph and drawing the line through the origin that is judged by eye to be the best fit to the points. On many occasions this method may be quite adequate, but when the points are scattered it becomes difficult to judge the position of the line, and some less subjective method, based on calculation, is advisable. Nevertheless, a graph should always be plotted, for a knowledge of the distribution of the points is necessary both in deciding what form of calculation to apply and in checking the fit of the calculated line.

It may readily be seen that the proportionality factor could be calculated as the average ratio of y to x or as the ratio of the average y to the average x, or in other ways. The type of calculation to be employed depends on whether all the points on the graph are to be given equal weight or whether some are to be considered to be more reliable than others. Here the concept of weight means that in calculations a point of weight 2 is made to have twice the effect that it would have if it were of weight 1. Now since the values of y are not necessarily equally accurate they should not necessarily be given equal weight. Since the values of x are not subject to error they have equal weight.

It may be shown by the principle of least squares[3] that the best line that is forced to pass through the origin has the equation $y = bx$, where $b = \Sigma wxy / \Sigma wx^2$. The sign Σ denotes summation over all the pairs of values of x and y, and w is the weight of the appropriate value of y. It is clear that the manner in which w varies for the different values of y will affect the value of b.

Now it may be further demonstrated that the correct weight to be given to any point is the reciprocal of the square of its standard error. Hence

$$b = \sum \frac{xy}{s^2} \Big/ \sum \frac{x^2}{s^2}$$

where s is the standard error of a given point. It is not necessary to know the exact value of the standard error of each point in order to

apply the above equation; it is sufficient if the form of its variation from point to point is known. For example, where the accuracy is constant over the range of measurement the equation becomes $b = \Sigma xy/\Sigma x^2$. If the standard error is proportional to the value of x, i.e. if the percentage accuracy is constant, $b = \Sigma(y/x)/n =$ average (y/x). If the standard error is proportional to the square root of x, $b = \Sigma y/\Sigma x =$ average y/average x. Many more cases are possible, especially if functions of experimental observations are plotted.

Although in practice the precise form of distribution of the errors may not be known, theoretical considerations and an examination of the experimental results will usually indicate roughly into which category the given graph falls and so enable a reasonably accurate estimate of the proportionality factor to be made. It is helpful to examine the graph closely for the distribution of error so that in the calculations too much weight is not given to points that appear to be subject to large errors. So-called freak results need to be watched carefully so that the proportionality factor is not unduly influenced by them.

The following examples illustrate some of the points that have been mentioned. Both concern the estimation of the percentage of carbon in a clay by different methods. Method A is an accurate reference method for the determination of carbon and is common to both examples. Method B in the first example and method C in the second example are approximate methods, both of which give values that are roughly proportional to the correct one. The proportionality factor had to be found empirically from the results given in Table II. In both examples the values obtained by Method A are denoted by x and those obtained by the alternative method by y.

The graph for the first example is shown in Fig. 3. There are no *a priori* reasons for making any assumption about the error distribution, but the graph seems to indicate that the error is roughly proportional to x. From the appropriate form, average (y/x), the factor is 0·770. This happens to agree exactly with the value obtained by other workers. Had the result from the unusual clay with a very high carbon content not been available the factor would have been 0·771—the difference is negligible. If the form appropriate to a constant error distribution had been used the factor would have been 0·780, which is not greatly different from the previous value. In this case, however, if the results from the unusual clay had been omitted the factor would have been 0·807, and the difference is appreciable.

In the second example one would expect the accuracy to be roughly constant, and the graph in Fig. 4 appears to confirm this. The expression $\Sigma xy/\Sigma x^2$ leads to a factor 1·458. The broken line on the graph corresponds to this and does not seem to be a very satisfactory fit at moderate values of the carbon content. It is clear that the result for the clay with the unusually high carbon content influences the value of the factor to a large extent; if this result is omitted the factor

TABLE II

Results for estimation of carbon content (%) of clays by three methods: method A direct, and methods B and C indirect

Method A	Method B	Method C	Method A	Method B	Method C
0·74	0·55	1·09	1·79	1·24	2·80
0·13	—	0·06	0·58	0·36	2·09
0·15	0·10	0·00	6·55	5·58	9·68
1·47	1·17	2·33	2·54	2·32	4·08
0·52	0·38	0·50	1·43	1·40	2·80
0·27	0·22	0·53	2·74	2·37	3·93
0·99	0·70	0·65	6·08	5·39	8·22
0·53	0·49	0·72	0·75	0·64	0·28
1·53	1·17	2·46	0·16	0·11	0·35
0·87	0·70	1·54	5·06	4·51	7·49
0·28	0·12	0·70	0·86	0·80	1·41
0·27	0·16	0·00	0·16	0·12	0·00
3·07	2·11	4·82	11·43	8·30	15·89
0·25	0·18	0·30	0·19	0·14	0·18
0·25	0·16	0·64	0·13	0·09	0·28
0·29	0·28	0·78	0·25	0·14	0·63
0·12	0·09	0·12	0·19	0·22	0·44
1·50	1·13	2·36	0·33	0·18	0·56
1·31	0·98	2·14	0·48	0·31	1·09
0·31	0·15	0·08	0·26	0·17	0·01
0·14	0·15	0·00	0·17	0·11	0·00
2·98	2·18	4·53	0·20	0·13	0·53
6·84	4·51	9·94	0·41	0·41	0·68
2·15	1·59	3·68	0·37	0·31	0·67
1·35	1·15	1·84	0·98	0·65	0·87
0·40	0·33	0·97	3·90	3·67	6·55
4·18	3·24	6·14	2·05	1·69	3·88
0·22	0·21	0·52	1·30	1·09	2·21
0·38	0·24	0·40	0·26	0·26	0·36
0·24	0·18	0·46	0·72	0·50	1·61

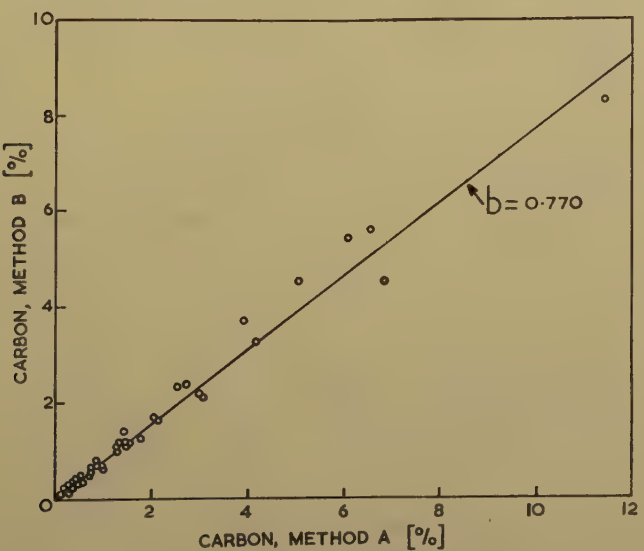


FIG. 3. Comparison between indirect method B and direct method A for determining the percentage of carbon in clay.

becomes 1·493. The full line on the graph corresponds to this factor and seems to be a somewhat better fit at moderate values. Since the relationship would be used mostly at low and moderate carbon contents and only very rarely be used at a high carbon content, the omission of the outlying result seems to be justifiable. It must be admitted, however, that the relationship is not very close and hence that Method C is hardly satisfactory. If a constant percentage accuracy had been

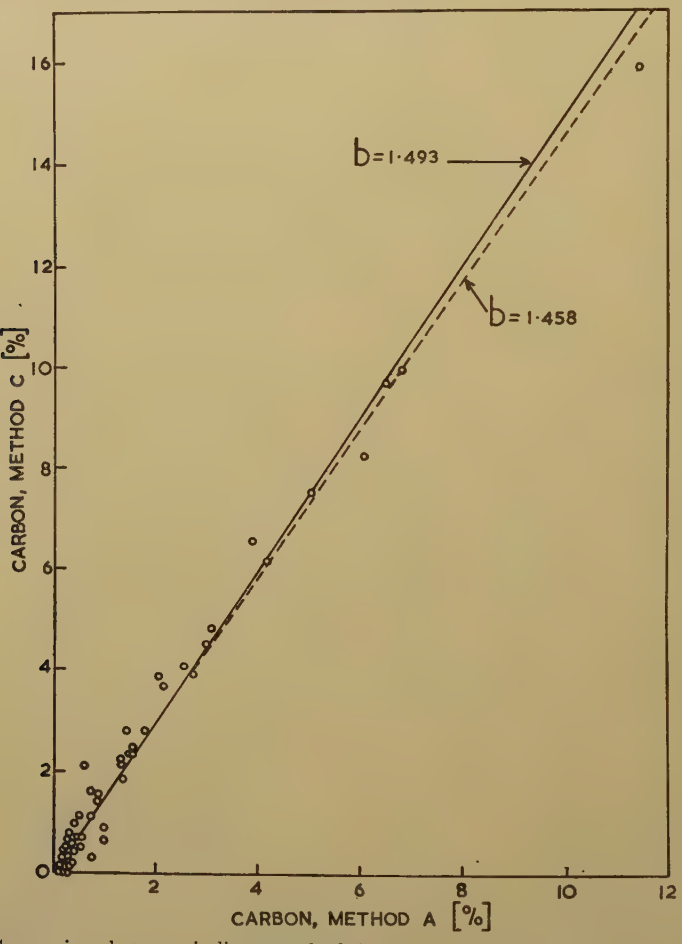


FIG. 4. Comparison between indirect method C and direct method A for determining the percentage of carbon in clay.

assumed, the expression: average (y/x) would have given a factor 1·499. In this case, of course, some of the values at low carbon contents would have exerted considerable influence; omission of the most extreme of them would lower the factor to 1·463. The third expression, average y/average x, would give a factor of 1·509.

The above examples illustrate the difference between some of the methods for calculating the proportionality factor. The difference will tend to be greatest when the relationship between x and y is not very

close. When the relationship is very close the three methods give practically identical values. Even more striking, perhaps, is the large effect that can be exerted by an outlying point; unfortunately such points are liable to occur where experimental material cannot be chosen but has to be accepted. In all matters relating to the interpretation of these relationships between variables it is important to give full weight to extra-statistical considerations, such as the manner in which the data were obtained and the purpose for which the results are to be used.

Estimation of the Importance of Factors

This section deals briefly with some points that arise when the correlation coefficient is used to estimate the importance of the factors influencing a dependent variable. It is important to realise that statistical methods in themselves cannot decide which are the factors governing the behaviour of a dependent variable. There must first be some cogent reason for deciding which may be the important ones in a given instance; the technique of correlation analysis may then be applied to test the hypothesis. With this proviso the correlation coefficient is a most useful tool.

The use of the word 'significant' is worthy of some comment. A correlation coefficient is statistically significant when the results giving rise to it are not likely to have happened by chance alone. The level at which a correlation coefficient becomes significant depends on the number of results available for its estimation; if there are sufficient results a correlation coefficient of only 0·3 may be highly significant. But in many circumstances a factor that showed such a low correlation would be regarded as of little importance.

The level above which a correlation coefficient is considered to be important depends on the point of view. In the simple case where only one assignable cause of the variation can be found and all the unaccounted variance, i.e. the scatter round the regression line, is thought to be due to random causes, a common tendency is to think that a correlation coefficient greater than 0·5 is indicative of an important factor. But if r is the correlation coefficient, the proportion of variance explained is r^2 and the proportion unexplained is $(1 - r^2)$. So unless r^2 is greater than 0·5 the cause which has been postulated is not the dominant one. In order to account for the greater part of the variation the correlation coefficient must therefore be greater than 0·71. That is not to say that smaller coefficients should be dismissed as of no account, but that the importance of a correlation should not be over-estimated just because it is statistically significant.

Care is needed when it appears that several factors may influence the dependent variable and that these factors are not completely independent of each other. The theoretical background should be considered first; when it appears that one of the possible causes may be merely acting through another it is necessary to calculate partial

correlation coefficients. The purpose of such coefficients is to measure the correlation due to a given factor when other known factors are kept constant. An example will be given in illustration.

The incidence of a certain type of mark in a ceramic plate was being investigated, and possible factors influencing it were thought to be the thermal expansion of the plate and the percentage of silica in the plate. Experimental investigations showed the following correlations:

number of marks per plate: expansion	+ 0·615
number of marks per plate: percentage of silica . .	+ 0·427

It was also known that the percentage of silica and the thermal expansion are usually correlated. The question then arose: does the high thermal expansion cause the mark and does the amount of silica only appear to cause it because it affects the thermal expansion, or is the percentage of silica the true cause of the mark, or do both expansion and the silica percentage operate ? Here the partial correlation coefficients give useful evidence although they cannot prove what is the cause.

In the present case the simple correlation coefficient of thermal expansion with silica percentage was + 0·663. The partial correlation coefficient of number of marks with silica percentage (expansion being constant) was + 0·032, an entirely insignificant quantity. The partial coefficient of number of marks with thermal expansion (silica percentage being constant) was + 0·491, which was statistically significant although not indicative of a predominant factor. Hence the silica percentage seemed to have no independent effect, a conclusion which was in line with other reasoning; for the marks were far more likely to have been due to a physical property of the material than to its chemical constitution.

Finally, a similar example is given showing the misleading impression that can be created by simple correlation coefficients. They are correct as statements of co-variation, but they may be misleading as indicators of causes. A factory process was being studied, namely the addition of clays and alkali to water to give a suspension, and subsequent filter-pressing to give a plastic cake of clay of roughly constant moisture content. The quantities measured from day to day were the specific gravity, *p*H value, and viscosity of the suspension, and the plasticity of the resultant cake of clay. The properties of the clays used are liable to vary somewhat from day to day, and the investigation was concerned primarily with process control. Table III gives the simple correlation coefficients obtained over a period.

It would be wrong to conclude that the specific gravity, *p*H, and viscosity of the suspension all influence the plasticity of the cake strongly and to a roughly equal extent. It is clear, however, that some correlation exists and that the properties of the suspension are to some degree interdependent. On general grounds one would not expect the specific gravity of the suspension, which merely measures the ratio of water to

clays, to influence the plastic properties of a cake of clay that has had most of the water removed to give a constant water content. When two variables are held constant the remaining partial correlations are:

plasticity: specific gravity .. + 0·07
plasticity: *p*H value .. .. — 0·28
plasticity: viscosity .. .. + 0·29

TABLE III

Simple correlation coefficients between properties of clay suspension and press cake

	Plasticity of Press Cake	*p*H of Suspension	Viscosity of Suspension
Specific gravity of suspension ..	+ 0·48	— 0·31	+ 0·76
*p*H of suspension	— 0·54	—	— 0·58
Viscosity of suspension	+ 0·63	— 0·58	—

Thus specific gravity was shown to be unimportant. Viscosity and *p*H values were of only small importance and the coefficients were hardly significant. The multiple correlation coefficient of plasticity on viscosity and *p*H value together was 0·67, and this value was significant. The value was not appreciably increased by including the specific gravity. Hence there is some evidence that variation of *p*H and viscosity accounts for part of the variation in plasticity. The alkalinity of the suspension, which is measured by the *p*H value, would be expected to influence the plasticity of the press cake, and hence may be termed a cause. It is reasonable to suppose that a suspension of high viscosity would tend to give a press cake of high plasticity, and hence the influence of viscosity in the multiple correlation is understandable. However, the viscosity of the suspension could hardly be termed a cause of the plasticity of the press cake; rather should we say that both are concomitant effects of some other causal factor.

Acknowledgment

The author wishes to thank Dr A. T. Green, OBE, Director of Research of the British Ceramic Research Association, for permission to publish this paper.

REFERENCES

[1] Jessop, W. N. (1952). 'One Line or Two?' *Applied Statistics*, **1**, 131.
[2] Lindley, D. V. (1947). 'Regression Lines and the Linear Functional Relationship,' *J. R. Statist. Soc. Supp.*, **9**, 218.
[3] Deming, W. E. (1943). *Statistical Adjustment of Data.* Wiley, New York.

ECONOMIC FORECASTING IN GREAT BRITAIN

ERIC SHANKLEMAN

After surveying early attempts to forecast the economic climate with the aid of 'business barometers' and estimates of the national income, Mr Shankleman discusses post war developments in forecasting and gives the results of some forecasts made by a number of economists for the Study Section of the Royal Statistical Society. Some forecasts recently made for 1953 are given.

THE ability to detect what lies ahead offers considerable reward in many spheres of life. Certainly it has always been a necessary, if not sufficient, quality for success in business. And today, more than ever before, capacity to forecast with reasonable accuracy is desirable in governments that have assumed responsibility for full employment programmes.

In many cases predictions will be based entirely on intuition, in others on carefully analysed information. In still other cases attempts will be made to apply in one way or another formulae that have seemed to be appropriate in the past. For many business purposes all that will be deemed necessary is to project into the future past trends, of sales figures for example. But for many this will need to be supplemented by some attempt to foretell the likely conditions to be expected in the market, in other words, the economic weather ahead.

It is with attempts to forecast the economic weather that we are primarily concerned in this article. We are concerned in particular with forecasts over short periods of up to a year, the sort of period in which business is most likely to be interested for budgetary purposes. To what extent have systematic methods been used in the past, how successful have they been, and what lesson do they offer present-day prophets ?

Business Barometers

Economic forecasting is largely a twentieth-century phenomenon. The main instrument in the early days was the 'business barometer,' which usually took the form of a chart comprising a set of indices whose movements, or more particularly whose turns, were considered to be indicative of impending economic change in general business conditions. The choice of indicators was based usually on extensive research into the cause and form of earlier business fluctuations; the underlying idea was that the historical patterns of the past would reproduce themselves in the future.

Probably one of the earliest of these barometers[1] was the one constructed in the middle of the nineteenth century (and used very successfully if we are to judge from the fortune he amassed) by the Parisian

doctor Clement Juglar, who forsook medicine for the study of economic fluctuations. But interest in business barometers developed extensively at the beginning of the present century when French and American economists were actively engaged in the search of suitable indicators. The French Government had appointed a committee of experts to investigate means of easing unemployment arising from periodic depressions, and this committee isolated systematic series which included the prices of raw materials, the volume of coal consumption, and the volume of exports; the movements of these series were found to precede general crises and were suitable for forecasting purposes.[2]

Likewise, sensitive indices were also being sought at about the same time by the Americans R. W. Babson and J. H. Brookmire, working independently. And just before the first world war the Harvard Committee on Economic Research produced the Harvard Barometer incorporating three indices which represented stock market speculation, business conditions, and the supply of money. The index of speculation, based on prices of industrial stocks and shares traded on the New York stock exchange and on New York bank clearings, was expected to change approximately four to six months earlier than business conditions, which were indicated by bank clearings outside New York City, and commodity prices. Similarly the index of business conditions would tend to change about six months before the money index, based on short-term discount rates.

The first barometer to be constructed in respect of British economic conditions appeared in the 'Reconstruction Supplements' of the *Manchester Guardian* during 1922.[3] This used the Sauerbeck index of wholesale prices as 'a fairly reliable barometer of the general trend in business conditions.' The barometer, or trade chart as it came to be known, was revised and elaborated by the London and Cambridge Economic Service, which had been formed to make available to British businessmen statistical data indicative of future business conditions. The chart appeared monthly in the *Bulletin* of the Service.

Like the Harvard Barometer, its London and Cambridge counterpart consisted of representative series indicating the factors speculation, business conditions, and money. The speculation factor was represented by an index of industrial securities based on the share prices of some twenty companies operating principally in the United Kingdom. These prices were believed to represent 'the most financial forecaster of change in trade conditions'[4] on the basis that prices of speculative securities indicated the market's expectation of profit, which in turn influenced general business activity. To represent business conditions against which the movements of speculation and the money index were to be compared two indices were used: the index of exports, in the absence of more comprehensive statistics of production, and the index of wholesale prices. To express movements in the monetary forces, an index was constructed of the average of bank rate, bankers deposit rates, and discount rates on short-dated bills. Low money rates,

the theory ran, offered incentives to business and the speculator to increase activity. During periods of higher rates production had diminished and stocks of commodities and goods had been reduced so that the stage had been set for industrial revival.

The London and Cambridge Economic Service was obviously intended to serve a pressing need; in November 1922 Sir William (now Lord) Beveridge, then Director of the London School of Economics, was reported in *The Times* to have drawn attention of Manchester businessmen to the trade chart being compiled by the London and Cambridge Economic Service and to have stated that 'the chart might be studied by the businessman with the same confidence with which he now consulted the indications of the thermometer or the barometer.'

The forecasts by the Service were included in the *Bulletin* in its monthly review of 'General Business Conditions.' It is to these, therefore, that we must look to assess what assistance was derived from movements in the four representative series included in the trade chart. Quite early on it was clear that by themselves they were not sufficient for forecasting purposes. Thus in July 1923 it was reported that apart from political events progress would be resumed in the Autumn but 'the political situation, which of course we cannot forecast, may prevent improvement.' The difficulty of forecasting political situations was to re-occur frequently. Another weakness of the chart as a basis for forecasting was to be noted soon after, for in September 1923 the *Bulletin* commented that while in normal conditions the principal series would have given good grounds of anticipated new confidence and business activity, special conditions obtained which made a trustworthy forecast impossible. Throughout the history of commercial cycles, the *Bulletin* recorded, special conditions of one type or another had occurred which delayed or altered the normal course of events.

It is not easy to answer the question of major interest to us in this paper, that is the extent to which the chart was of value in preparing the forecasts of the Service. Certainly references to sequential movements of the indices in the trade chart were few; after the early issues one has to wait until June 1934 for such a reference. Most forecasts on the other hand drew attention to significant trends in the many other series included in the *Bulletin*.

There were clearly deficiencies in the extent to which the indices of prices and exports accurately represented the current state of business activity. Production can increase when prices fall, indeed a fall in prices may be due to increased production, costs declining as a result. Again the export index would not represent equally well all sectors of industry; further the index was one of value and not volume and consequently reflected price as well as volume changes.

But, even this apart and assuming that the four series comprising the chart had moved in what had been hitherto considered a normal sequence, the considerable difficulty still remained of determining whether or not the leading index of industrial securities had reached its

peak or trough. Yet it is in their ability to warn of *radical* changes or turns in economic conditions that barometers are likely to be of greatest use to business. Told that conditions are likely to improve or deteriorate slightly the businessman may find little cause to alter an existing programme; at most, what would be called for is its acceleration or retardation. But the warning of a distinct turn in the direction of business activity may suggest the complete reorientation of policy calling for serious expansion or contraction in such items as the size of labour force, supplies of materials, planning, and advertising appropriations. The cost of failing to comply with such a forecast or of adopting a faulty forecast are manifestly considerable.

The London and Cambridge Trade Chart was not the only barometer published in the pre-war period. In October 1923 the Federation of British Industries began the quarterly publication of their own barometer constructed 'on the basis that experience shows that trade is closely related to the speed with which money flows to and fro between the banks and trade, and that all changes in trade are *preceded* by corresponding changes in monetary conditions.'[5]

The barometer was published quarterly together with forecasts, an example of which is the following from the issue of the *FBI Bulletin* of 29th January 1953: 'The indications point to a marked improvement in trade during the next few months, though it is not possible to predict how far this movement will go on until more information is available regarding the future monetary policy of the Bank of England and the Treasury.'

'It is unfortunately necessary to qualify the above forecast by stating that its fulfilment is conditional on the absence of industrial disturbances due to strikes or similar causes.'

As in the case of the London and Cambridge barometer, it is not possible to detect from the forecasts the degree of reliance placed upon the trend in monetary indices, the major indicator, in preparing the forecasts published. Up to 1929 issues of the Economic Supplement of the Federation of British Industries which included the business barometer were still stating that 'by watching the movements of the money market and studying the actions of the bank in relation to current business conditions, it is possible to forecast ahead with a fair degree of accuracy the full course of trades.' After that period comments on monetary factors continued but, as was the case from the inception of the barometer, other factors attracted considerable attention. Indeed, their theories, the Federation stated, 'must not be taken as implying that the only influence determining trade-cycle movements is money, for there are instances of fairly definite expansions and contractions of trade unaccompanied by any very marked alteration in monetary conditions. There can, however, be no doubt that money was one of the outstanding influences in the major pre-war international fluctuations.'

From our examination of the two barometers described, it would not

seem unfair to say that the initial hopes, that forecasts of the future could be based to any large extent on the expected repetition of past patterns of economic behaviour, were not fulfilled.

Relationships important under one set of conditions may be superseded. The series used in barometers were taken to be symptomatic of the major factors causing change in economic conditions. But as soon as other factors become important or some existing factors lose importance, the value of the barometer diminishes. The problem is to know when such changes have occurred, for by their very nature one needs to watch all factors over considerable periods before their significance as indicators of economic change can be assessed.

National Income Forecasts

The business barometer relied on indirect and partial indications of activity in the economy. Had it been possible to examine the aggregate of business activity then the forces influencing the component parts of the aggregate might have emerged more clearly. In particular the important part that spending plays in determining the pulse of business activity would probably have been more widely examined. Two aggregative measures of business activity available in the 1930's were the Indices of Production prepared separately by the London and Cambridge Economic Service and the Board of Trade. More comprehensive was the Index of Business Activity compiled by the *Economist* newspaper. This aimed to take into account not only the rate of industrial production but also the tempo of distribution and other sectors of the economy. The object was 'to indicate the fluctuations in economic activity—that is the income-producing activity—of the community as a whole.'[6]

The most comprehensive measure of total business activity, however, is the size of national income. Commencing in 1934 a series of estimates on the national income were prepared by Mr Colin Clark for the commercial barometer issued by Messrs Pritchard, Wood, and Partners[7]; examples of these are given in Table I.

TABLE I

Estimates of national income for 1937

Quarter	1st	2nd	3rd	4th
		£ millions		
Wages and Salaries	560	573	582	587
Other Income	829	859	861	891
National Income	1 389	1 432	1 443	1 478
Consumption	1 182	1 190	1 180	1 217
Investment	198	209	211	208
National Income	1 380	1 399	1 391	1 425

These estimates represented *ex post* forecasts of what spending power had been in the immediate past (rather like a football supporter forecasting the result of a match that had already ended but whose results he did not know) and not an attempt to forecast national income in a period ahead. But in June 1937 Mr Clark prepared an *ex ante* (or 'before the event') forecast of the national income which was published in the *Economic Journal* of June 1937.[8] In this forecast Mr Clark sought to predict the size of the national income in the last two quarters in 1937.

This forecast is notable as being one of the earliest attempts on record to predict economic conditions in Great Britain by means of a model, albeit simple, incorporating important features of the economic system. A later example made in 1944 was by Mr Kaldor's exercise incorporated in Lord Beveridge's volume *Full Employment in a Free Society*.[9]

Economic events for a year ahead, Mr Clark suggested, were determined by what had already happened. Thus current investment plans at any particular moment in time had already been made on the basis of existing factors and unless price changes were very great no changes in those plans were to be expected. On the basis of past experience, Mr Clark assumed that investment twelve months ahead could be predicted from the following ratio computed in terms of current costs:

$$\frac{\text{profit per unit of output}}{\text{price of capital equipment} \times \text{rate of interest}}$$

Modifications were introduced in respect of certain industries; thus for shipbuilding, an 'incentive to invest' ratio could be more appropriately constructed from an index of freight rates and the price o new tonnage. To this sub-total for investment was added the value of armament work, and the changes in volume of working capital and the balance of payments.

To estimate the level of consumption Mr Clark used the past relationship between changes in national income and changes in investment, a relationship known to the economist as the multiplier, together with a lagged relationship between income and expenditure on the part of the rich. Checks of consistency in the exercise were also introduced. Thus he examined the extent to which increases in future output could be increased only at the expense of large increases in costs. If costs increased, he suggested, then 'however strong the incentive to invest and however precise the multiplier the advance of national income will be checked.'[8]

From his computations Mr Clark concluded that the national income would increase by some £84m. in the second half of 1937 as compared with the first half. His *ex post* estimates, subsequently published in the Pritchard Wood Commercial Barometer suggested the actual figure as being £36m. The direction of his forecast was right; national income was expected to increase and it did so, although he overestimated the size of the increase. In the course of his exercise Mr Clark also

produced a forecast of the Board of Trade Index of Industrial Production for the same period, which he suggested would be some five per cent. greater at the end of the year than it was during the first half of the year. The final index as published by the Board of Trade turned out to show a rather smaller increase of approximately four per cent. But these errors did not defeat the main purpose of Mr Clark's exercise which was to warn that a slump was foreseeable, beginning in 1938, and that the government should prepare plans for public works and stimulation of private investment and consumption. That his prognosis was correct there seems to be little doubt.

It would thus seem that this early attempt at the use of an economic model proved to be more successful than had been the earlier adoption of barometers. But there are points of resemblance; even though the economic model method allows for the building-in of checks on forecasts, both rely on past relationships to hold good in the future. But how stable are these relationships under changing economic conditions? Frequently insufficient experience is available to the statistician to study the degree of stability achieved. Further, any such relationships established from experience over a long term may be found to be quite deficient for making a short-term forecast. It was this deficiency which led in part to erroneous forecasts of national income and the level of employment by American economists following the end of the Second World War.[10]

Post War Forecasting

With the adoption by the Government in 1944 of a full employment policy, forecasts of future national income assumed a new importance. These forecasts formed part, at least until 1952, of the Economic Surveys published annually by the Government. Likewise in the United States of America, the Employment Act of 1946 (a similar 'full-employment' measure) requires the President to transmit biannually an economic report.

Apart from their importance in the planning of the national economy, the forecasts were undoubtedly of value to British business as representing the official view of economic prospects during the year. How accurate they proved to be is difficult to judge precisely. While many of these forecasts had the quality of being quantitative and therefore, *prima facie*, more easily checked than the judgments based on barometers, in practice many difficulties occur. The most important is that the national income forecasts were themselves based on estimates which tended to be revised in later publications. Definitions too may be amended. These must be borne in mind in considering Table II, which compares the forecasts and actual values, or more correctly *ex post* estimates, of gross domestic product at factor costs, personal consumption, and government expenditure at market price for the period 1948 to 1951.

TABLE II

Economic Survey, accuracy of forecasts

	Gross Domestic Product		Personal Consumption		Government Expenditure	
	Forecast	*Actual*	*Forecast*	*Actual*	*Forecast*	*Actual*
			£ million			
1948	9 700	10 500	7 675	8 004	2 050	1 914
1949	10 900	11 151	8 200	8 402	2 040	2 037
1950	11 630	11 896	8 640	9 041	2 220	2 025
1951	12 880*	12 414	9 675	9 970	2 510	2 440

*Including net income from abroad.

The table shows that in three of the four years the prospective *value* of the domestic product was underestimated. Errors could be due partly to predictions of the volume of output in the economy, and partly to predictions of the rise in prices which has been continual throughout the period.

A more precise comparison is possible for forecasts of industrial production and the number of unemployed. Forecasts of the former were included in Economic Surveys for 1949 to 1951; in 1949 and 1950 they proved to be much too low, whilst the figure for 1951 was too high. As regards unemployment in three years the number was over-estimated, while in 1949 and 1951 the estimates tended to be too low.

Techniques of National Income Predictions

We have already referred to the forecasting exercise by Mr Colin Clark in 1937 which depended partly on a simple system of equations. During the last fifteen years considerable advances have been made in construction of these systems or econometric models. Reference has already been made to the interesting example of a model used to examine possible post-war economic conditions which was included as Appendix C in Sir William Beveridge's *Full Employment in a Free Society* published in 1944, and examined in some detail by Stone and Jackson in the *Economic Journal* for 1946.[11]

In model building, the economist constructs a series of equations which represent how chosen variables interact to determine the behaviour of the economic system. The values of the constants of the equations are found by regression analysis applied to past data.

Little appears to have been published as to the methods used in making the official national income forecasts published in the Economic Surveys. An idea of the problems involved in making such forecasts may be derived from an approach outlined by Mr J. Downie in a paper on Economic Forecasting read to the Study Section of the Royal Statistical Society in November 1950.[12] It is Mr Downie's approach that we describe in the following paragraphs.

Starting with the assumption that full employment of the labour force, suitably defined, would be maintained as part of government policy, his method would involve estimating the level and composition

of the gross national product, the distribution of national income, and a consideration of the demand conditions to be satisfied if the projected level of gross national product were to be achieved.

The movement in *real* national product may be estimated by applying to projections of the numbers employed in different industries the expected value of output a head in each. If we call the latter a_i and the numbers employed in an industry n_i then the *real* value (i.e. in terms of money of constant value) of the contribution of industry to gross national product Y would be $\Sigma n_i a_i$. As these factors are interrelated an error in the output estimate may lead to errors in the forecast of labour distribution. But these are not likely to be very large over the short periods under consideration.

The data required are most readily available for those industries included in the Index of Industrial Production.[13] Estimates of the output of services, both public and private and including the armed forces, may be made by assuming constant productivity a head. The anticipated value of agricultural output needs to be added as does the increase in rental value of houses. Agricultural output is estimated from data available on areas under cultivation and livestock population, and the increase in rental value of houses is based mainly on the expected increase in the stock of houses. The computation involved may be demonstrated by reference to Table III which shows how an estimate of gross national product might be prepared.

TABLE III*

Hypothetical estimate of future gross national product by industry of origin, at current factor costs

	Output a Head (£)	Average Numbers Employed ('000)	Gross Product (£m.)
Agriculture, etc.	544	1 139	619
Mining and quarrying	476	856	408
Manufacturing	548	8 736	4 789
Building and contracting	494	1 449	716
Gas, water and electricity	669	362	242
Transport and communication	559	1 754	981
Distributive trades	576	2 600	1 497
Other services	274	3 968	1 090
Total production and trade	496	20 864	10 342
Public services and defence	521	2 177	1 135
Ownership of dwellings	—	—	238
Domestic services to household	—	—	92
Services to private non-profit-making bodies	—	—	58
Gross domestic product	—	23 041	11 865

* The total of gross domestic product shown in the table differs from gross national product in that it excludes net income from abroad.

The next stage is to allow for any changes in factor costs consisting of wage and salary costs, and gross profit margins. This introduces

'an empirical question on which one has to make the best guess possible in any particular case on the basis of whatever information is available of past movements . . . and whatever one knows about likely movements . . .' As with other techniques, the accuracy achieved depends to a large extent on the accuracy of individual judgments. Less respectful writers might call these 'guesses'; and cynics might be disposed to agree with the late Sir Hubert Henderson that in the process of guessing key factors in such forecasts 'a sort of Gresham's law is apt to operate under which wishful thinking drives out common sense.'[14]

Study Section Survey of Economists' Opinions

If the interwar period was an era of the business barometer, the post 1945 period is likely to become the era of opinion polls. But however one may value the results of asking people how they would vote in a forthcoming election, there would appear to be some merit in surveying the opinions of economists in industry and academic life as to the prospects for economic conditions, if only to examine how their views differ from the official outlook. The committee of the Study Section of the Royal Statistical Society, when organizing its first survey in 1951, had in mind the additional purpose of stimulating discussion of the problems involved in economic forecasting.

Questionnaires were sent to sixty-six economists in the universities, industry, commerce, banking, and the press, inviting them to predict the movement during 1952 of the following five economic indices:

(i) The Central Statistical Office's Interim Index of Industrial Production—the 'All Items Index,' average for calendar year 1952.

(ii) Index of Volume of Exports, average for calendar year 1952.

(iii) Unemployed as percentage of number of employees, average for calendar year 1952.

(iv) Personal expenditure on consumers' goods and services (at average 1948 prices). Percentage change, 1952 compared with 1951.

(v) Interim Index of Retail Prices—the 'All Items Index' percentage change, 1952 compared with 1951.

The sample was chosen, with no pretence at randomness, from among economists and statisticians who were believed either to have been active in business forecasting on behalf of the firms by whom they were employed, or interested in economic forecasting. It soon become obvious from correspondence that in a small number of cases, predominantly among academic economists, some had been included who lacked the requisite interest. The composition of the sample chosen and the sample responding is shown in Table IV.

To assist economists to complete the questionnaire they were provided with tables showing in some detail the trend of the five indices over the last four to five years. This was a procedure found to be of value in similar surveys organised by the Chicago Chapter of the American Statistical Society.[15]

TABLE IV

Numbers of economists (a) *invited and* (b) *completing questionnaires*

	Numbers		Percentage, Completing: Invited	Composition of Active Sample
	Invited	Completing		
Universities	17	10	59	22%
Press, banking and commerce ..	21	14	66	31%
Industry	26	21	81	47%
Total ..	64	45	70	100%

Many of the economists who refrained from completing the questionnaire commented on the experiment in correspondence. At least one academic economist was frankly shocked that anything of the sort should have been attempted, and wrote: 'I did not realise the Royal Statistical Society encouraged astrological economic statistics.' Another considered that as the future depended so much on the terms of trade and government policy, both of which he was unable to predict, he could not complete the questionnaire. Other reasons why economists failed to co-operate were lack of experience in economic forecasting and the absence of information as to the object of the survey. One economist in industry wrote: 'In my view forecasts of this kind are accepted by far too many people as more accurate than in fact they can hope to be, and unless there is some special reason for making an attempt, I think the exercise is best not undertaken.'

Many of those responding stressed the wide margins of error to which they considered their forecasts were subject and this point featured prominently in the discussion on the results at the Study Section's meeting in May 1952.

Choice of Indices

The choice of indices to be forecast was made mainly on grounds of their interest to industry or because of the importance of the part they played in determining the size of the national income. The expected level of industrial production, as we have already seen, is an important determinant of the volume of resources available for use in the economy generally.

In an economy entirely dependent as to whether it pays its way or not upon the oversea sales of its products, the expected level of exports is of obvious interest. And like the unemployment figure, it had topical interest in making available in quantitative terms the views of those engaged in industry.

Personal consumption was considered to be of direct interest to those in industry attempting to forecast the sales of various types of goods and services. One approach used is first to estimate, on the basis of past experience (amended, if necessary, in view of the likely effect on sales

of the anticipated change in the overall volume of personal consumption and other relevant factors) the proportion of total personal expenditure on consumer goods and services attributable to the industry concerned. This proportion is subsequently applied to the forecast total of personal expenditure to give an estimate of the industry's sales. The final stage is for the individual firm to estimate its own proportion of the industry's sales and so the volume of its sales turnover in the commodity concerned.[16]

The predicted value of the Retail Price Index is of interest as indicative of the measure of inflation expected in the economy; but it is also of interest to many firms, the wages of whose employees are either statutorily or otherwise linked to movements of the index.

The Results of the Survey

Probably the most interesting feature of the results was the close similarity between the mean forecasts of the university and other economists. Only for predictions of changes in the index of personal expenditure on consumer goods and services was there much difference between the two forecasts, but even this is scarcely statistically significant.

Both groups anticipated only a small increase in the level of industrial production, hardly sufficient to achieve the Chancellor of the Exchequer's hoped-for increase of £250m. in the volume of total output in 1952. An increase of this order would have required an increase in the Index of Production of approximately 4 per cent.—a level which few economists in the sample were prepared to expect.

Neither was there very much confidence in the achievement of the Chancellor's hope of an increase in the volume of exports; most forecasters clearly expected a fall.

The number unemployed as a percentage of insured employees had averaged 1·6 per cent., 1·5 per cent., and 1·2 per cent. in the years 1949, 1950, and 1951. A relatively large increase to 2 per cent. was expected in 1952; at the time of making the survey however the monthly average stood at 1·9 so that a considerable change was not expected over the remaining months of the year.

Of the 45 respondents to the survey only eight were prepared to forecast increases (mainly of one or two per cent.) in the level of personal consumption in 1952 compared with 1951; seven forecasted no change. An examination of the dispersion of the forecasts indicated a wide variation in opinions as to the extent of the fall in personal consumption expected among the remainder of the forecasters.

Academic economists co-operating in the survey in the main expected the level of retail prices to increase by 8 per cent. to 10 per cent. in the year. The average of their predicted increases equalled those expected by non-university economists. But in the latter case the dispersion of forecasts was much greater. In two instances, it is interesting to note, falls in the index of prices were predicted.

The averages of all forecasts together with the values of the indices achieved in 1952 are given in Table V. It will be seen that the forecasts were reasonably accurate for the levels of unemployment, retail prices, and personal consumption. The largest errors were made in predicting industrial production and the volume of exports, although these errors were not as large as were those in official forecasts. These last two indices are inter-related; a redeeming feature therefore is that they indicate some internal consistency between forecasts. This was evidenced by the correlation coefficient of 0·43 between them. Likewise a coefficient of 0·41 was obtained between the indices of industrial production and personal consumption. A coefficient of roughly the same size was obtained between personal expenditure and unemployment.

TABLE V

Comparison of predicted indices with results for 1952

	Predicted	Actual
(i) Index of Industrial Production percentage change on 1951	+ 1·1	− 3·0
(ii) Index of Volume of Exports percentage change on 1951	− 1·6	− 6·2
(iii) Unemployed as percentage of number of insured persons	2·0	2·0
(iv) Personal Expenditure on Consumers' Goods and Services (at 1948 Prices) percentage change on 1951	− 1·6	− 1·1
(v) Interim Index of Retail Prices..	135	136

In February 1953 a second enquiry was carried out relating to the current year. Forecasts were again invited in respect of the indices used last year. The size of the sample however was increased and respondents were asked in addition to state those factors which they considered could most likely effect the accuracy of their forecasts.

Of the 109 economists invited 59 have completed questionnaires, or some 55 per cent., as against 70 per cent. in 1952. The results of the survey are given in Table VI.

Conclusions

However difficult good forecasting may be, it has to continue. The lesson of our review of techniques would seem to show, however, that there are no short cuts. What may be appropriate indicators of economic change at one time may quite shortly afterwards prove to be of little importance, and only as a result of continual analysis of conditions can the value of different indicators be assessed. In such circumstances the value of barometers has proved small. Nevertheless research into their behaviour continues and an extensive study in America

TABLE VI

Level of indices in 1952 *and predictions for* 1953

	Actual 1952	Average Prediction 1953
(i) Index of Industrial Production	114	117
(ii) Index of Volume of Exports	95	98
(iii) Unemployed as percentage of number of insured persons	2	2
(iv) Personal Expenditure on Consumers' Goods and Services (at 1948 Prices) percentage change on 1951	− 1·1	+ 0·75
(v) Interim Index of Retail Prices..	136	140

over a long period has suggested that providing sufficient numbers of appropriate series are used it may be possible by suitably analysing the movement within the group of series to obtain a reliable guide as to turns in the progress of the trade cycle.[17]

No similar research has been reported in this country but the subject raises the question whether, with the advent of the full employment and similar policies, cyclical fluctuations will be damped in future. Will the redistribution of incomes that has occurred in our society, together with social insurance schemes and the increased strength of the trade union movement, ensure that 'under-consumption' will no longer be an important cause of depression ? Is it that our economy might prove too inflexible to compete successfully in all important export markets ? Clearly all these factors may prove important in influencing the course of economic events; again, cycles in the past have proved international and it is difficult to see how Great Britain could be completely insulated, if at all, from conditions overseas, except at a most undesirable cost.

By its ability to embrace many of such factors the national income exercise for forecasting purposes offers attractions. And by enabling the forecaster to alter quantitative values of assumptions, the implications of alternative movements in the economy may be assessed. The value of such exercises could be enhanced however by the availability of more up-to-date information, in particular quarterly estimates of national income instead of the annual figures at present officially published. Further, much more might usefully be done in Great Britain to survey regularly consumers' spending and the plans of industry as to investment.

REFERENCES

[1] Hutchinson, T. W. (1952). *A Review of Economic Doctrines*, 1870–1929. Oxford University Press.

[2] Pavlovsky, G. A. (1922). 'Business Forecasting a Statistical Problem,' *Tracts*, Royal Statistical Society, Series 5, **134**.

[3] Snow, E. C. (1923). 'Trade Forecasting and Prices,' *J. R. Statist. Soc.*, **86**, 332.
[4] London & Cambridge Economic Service (1923). *The Bulletin*, **1**, 1.
[5] Federation of British Industries (1923). *The Bulletin*, **6**, 629.
[6] Crowther, G. (1934). 'The Economist Index of Business Activity,' *J. R. Statist. Soc.*, **97**, 241.
[7] Pritchard, Wood, and Partners (1934). *A Commercial Barometer*. **1**, 1.
[8] Clark, Colin (1937). 'National Income at its Climax,' *Econ. J.*, **47**, 308.
[9] Beveridge, W. H. (1944). *Full Employment in a Free Society*, Allen & Unwin, London.
[10] Sapir, Michael (1948). 'Review of Economic Forecasts,' *Studies in Income and Wealth*, **11**. National Bureau of Economic Research, New York.
[11] Stone, Richard, and Jackson, E. F. (1946). 'Economic Models with special Reference to Mr Kaldor's System,' *Econ. J.*, **56**, 554.
[12] Downie, J. (1951). 'Forecasting National Income,' unpublished but available in the library of the Royal Statistical Society.
[13] HMSO (1952). *An Index of Industrial Production.*
[14] Henderson, Sir Hubert (1947). *The Uses and Abuses of Economic Planning*, Cambridge University Press.
[15] American Statistical Society (1951). *The American Statistician*, **5**, 16.
[16] Friday, Frank A. (1952). 'The Problem of Business Forecasting,' *J. Indust. Econ.*, **1**, 55.
[17] Moore, G. H. (1950). *Statistical Indicators of Cyclical Revivals and Recessions*. Occasional Paper No. 31. National Bureau of Economic Research, New York.

STATISTICS AND PREDICTION

CYRIL DAVENPORT HUGHES

Mr Hughes writes from the point of view of an organisation manufacturing components used on industrial equipment and requiring to predict future demand for the purposes of framing policy. He examines statistical techniques based on regression analysis and the analysis of time-series and finds them wanting. He concludes that predictions must be based on a considerable amount of reliable, detailed information about a firm's customers and what happens to its products.

Henry Ford once said 'History is bunk.' There is today a sense in which 'Prediction is bunk.' To say this is not to underrate the importance of forecasting—without forecasting there can be no forward planning, and all business enterprises depend for their very existence on some forward view of sales prospects, however unsophisticated their terms of definition. There are, however, many fields in which consistently accurate forecasting is unattainable, and in attempting to eliminate errors and make forecasting more 'scientific' the statistician may soon realise that knowledge and judgment can be much more important than acquaintance with statistical techniques. No fancy formulae can compensate for inadequate basic data.

Let us look for a moment at the background to business forecasting in a large organisation manufacturing, say, a wide range of components in use on industrial equipment in the home and export markets. Since 1939 or thereabouts effective demand for such products has probably been in excess of productive capacity. There are available, therefore, no reliable statistics whatever on demand trends over this period. Actual sales will represent merely the curve of production. Demand, as represented by the order book, will have been inflated by anticipatory and speculative ordering: inability to meet demand from certain manufacturers of industrial machinery may have led to recourse to competitors' products and maybe to permanent loss of markets, without any immediately obvious reflection in the balance of orders. If data over the period 1939–51 are unsuitable as a basis for prediction, the data reflecting the recent fall in demand ushered in by the recession in consumer goods, and the serious restrictions placed on Commonwealth and foreign imports in an attempt to correct a crisis in the balance of payments, are equally unsuitable for prediction. General experience shows that cancellation of orders under such circumstances reflects not only the elimination of speculative ordering but also willingness to draw on stocks and postpone ordering until the pipelines are cleared. The latest statistics will, moreover, show the re-emergence of seasonal cycles in demand in addition to the main downturn of the trade cycle and are

unlikely to be of much use for prediction. Under such circumstances talk of 'stability of data,' factors 'under control,' and 'homogeneity of material' is largely academic. The basic data are not available.

In addition there is the further difficulty of the distinction between manufacturing demand and replacement demand. Unless there is a clear distinction in the internal statistics of industrial organisations between sales of products for incorporation in new equipment and sales for replacement of components in old equipment it may be difficult to prepare the basic data required for what is soon to become a most important element in business forecasting. I refer here to the volume of demand for component replacement likely to result from the tremendous quantities of original equipment produced in this country since the war. During the period 1946–52 British industry produced £270m. of agricultural machinery (excluding tractors), £420m. of textile machinery, £260m. of machine tools, £190m. of mechanical handling equipment, etc. Before very long the statistician will have to look into his crystal ball very intently if he is to give much help on this problem.

Another limitation on the statistical approach to forecasting is that to some extent the forecast is itself an independent variable which one may have to take into account in assessing likely sales. This is not quite as silly as it sounds. Sales campaigns take time to mature, they cannot be mounted at a moment's notice. Where existing demand falls only a little short of productive capacity a major sales campaign might lead to considerable embarrassment. On the other hand, adequate warning of a considerable surplus of capacity over demand may make it possible to stimulate demand to a level in excess of that of capacity—if the situation merits opening up new fields of application. Accordingly, volume of sales in the future will depend in part on adequate knowledge of future production and the gearing of sales campaigns.

Statistical Techniques

How far can the statistician help in these circumstances ? What are the techniques available, and how can they be applied ? Perhaps the most familiar is that of multiple regression analysis, in which the statistician looks for an equation in which effective demand is a function of a number of causal variables. It is my experience that for durable-use producer goods the relationship between causes and effect is highly complex, and one for which statistical series adequately reflecting changes in certain of the causal variables are generally not available. Assumptions of linear relationships may, moreover, involve deviations between computed and actual values of the dependent variables so great as to nullify the usefulness of this approach. Technical changes, the effect of recourse to competitors, changes in price of competing products, and changes in demand for consumer goods which in turn affect demand for capital goods, are difficult, if not impossible, to state statistically.

To introduce time as a causal variable introduces as many problems as it solves if the data relate to a historical period with characteristics

different from those of today, or from those which we expect for tomorrow. In any case, demand for new products cannot be forecast by this method, and for an established product the data since 1939 are likely to be affected by conditions which we outlined above. Data prior to 1939 may in turn have little relevance to the facts of today.

Another problem is the degree of correlation between causal variables themselves—here we may be in the field of economics, where the existing knowledge of the relationship between demand for durable-use producer goods and other factors in an economic situation gives little cause for confident prediction. It may also be necessary to take into account the time-lag between series of data which seem to be related to a particular causal variable.

Assuming that despite these problems we arrive at our equation, in order to use it we must now forecast each of the causal variables. If this is to be done by multiple regression analysis we presumably go through the above process once more for each of the variables, on the basis of probably even more unreliable data. Unless we pursue this process to the point of exhaustion we decide at some stage to 'think of a number' or, more politely, we make a subjective or qualitative forecast (i.e. a guess).

Some statisticians encourage the use of relationships between national statistical series such as the national income, personal disposable income, savings, and demand for particular products. Theoretically this approach has the advantage of relative stability in the basic data. In practice, whatever the virtue of such methods for population prediction, they often have little relevance to forecasts of demand for particular products, which may not in fact be highly correlated with changes in the general level of economic activity. In any case the official statistics (e.g. of personal savings) often show revisions of such magnitude that they are quite useless for this purpose. Where such figures are used it is noteworthy that 'plans' usually fade into 'targets.' Where statistical series bear a margin of error which is sometimes even greater than their actual fluctuations there can be little virtue in using them.

Practical experience in forecasting shows that multiple regression analysis under all but ideal circumstances is only useful where one is dealing with generalised trends of demand over fairly long periods, chiefly as a rough check on estimates made by other means and as a method of calculating the relative importance of causal variables over a historical period.

Another technique is to use cyclical fluctuations or oscillations. It is true that adequate time-series analysis, showing for each product of an organisation a historical analysis of random fluctuations, seasonal cycles, trade cycles, and the relationship of each to basic trend, gives much useful information about the differences in the pattern of movement as between different groups of products. But it is very doubtful whether this method is of much use for prediction. As long as demand exceeds supply the sales curve will reflect the production curve—in

other words, the basic data are once again not available; moreover, when the downturn in demand appears, any emergent season cycle may be swamped, or alternatively greatly exaggerated, and the cyclical fluctuations may be difficult to disentangle. Even where adequate data are available there is rarely any strict recurrence of cycles, they vary in magnitude and timing; and projections, either freehand or mathematical, are very hazardous.

Information required for Prediction

If the mathematical techniques available to the statistician are of such doubtful value for prediction what techniques are available for this purpose? The only really satisfactory method appears to be to construct estimates of future demand after first clearing the ground by attempting to answer a series of questions concerned with the product. The sort of questions involved will be:

(i) What are the general features of the product and how accurately can usage be divided up among different types of consumers?

(ii) What is the company's selling and price policy for the product?

(iii) Who are the customers and ultimate consumers? What do they want? What will they want?

(iv) What influences govern actual as distinct from potential demand? Which of these influences are under our control? To what extent are we dependent on influences exerted on other industries at first, second, or even third remove from our own?

(v) What is the life of the product? Does it vary according to working conditions?

(vi) What is the extent of the goodwill?

(vii) Is demand exercised directly for component replacements and indirectly via complete industrial equipment? If the former, what are the prospects for the products of the industry of which the machinery incorporates our product? If the latter, what are the prospects for sales of new equipment which incorporates our product?

(viii) Is there competition? If there is, what is its basis—price, delivery, service, or quality? Or is it based on a dissimilar product serving the same purpose?

(ix) To what extent is competition, and demand itself, affected by trade agreements, quotas, tariffs, import licensing, etc.?

The first step must be product and market research, in which an even more comprehensive list of questions receive detailed treatment. One of the results of this questioning will be to establish the relative importance of the general economic weather to the prospects for a particular product. Where this is important the next step will be an economic analysis, showing the relevance of past, current, and estimated future trends to demand for the product. This analysis should then be the subject of rigorous treatment and detailed comment by executives as a group. Given a generally acceptable economic analysis we come to the real crux of the problem. Without a highly efficient, up-to-date

information service none of the subsequent work will be worth the paper it is written on. We must know enough about all our main customers to be able to judge the trend of their manufacturing programmes, to judge whether or not such programmes as they announce are likely to be fulfilled, to judge objectively our own prospects in the market, to judge the influence of economic forces on the timing, direction, and level of replacement demand, to judge prospects for new applications of existing products, etc. At this point adequate information, some of which will indeed be in statistical form, and sound judgment are much more important than involved statistical techniques.

In general it may be said that there is nothing new in all this. True, without this approach no forecasts can have much validity. But it is the experience of many in industry that the statistician often embarks on elaborate statistical calculations and projections at a moment when time and energy could most profitably be spent on developing a first-rate information service to provide essential background material.

Once estimates have been made in quantitative terms on the basis of known applications and the prospects of new applications, aggregates can be roughly checked by means of multiple regression analysis; but the basic estimates will have been made on the foundation of factors which can be clearly stated and graded according to degree of expected reliability.

Prediction of Indexes of Wage-rates and Machinery Prices

A related problem facing businessmen today concerns the extent to which Inland Revenue wear and tear allowances (based on original cost) understate true depreciation (based on replacement cost). Some companies use indices of machinery prices, others indices of wage-rates as being appropriate and relatively stable measures of changes in the value of money rather than calculate their own indexes of replacement costs. This is not the place to discuss the relative merits of the different approaches, but forward planning now requires long-term projections of these index numbers. An important distinction must be made here—the statistician or economist is here trying to estimate the movements not of wage-rates or machinery prices *per se*, but of a particular index of wage-rates and a particular index of machinery prices. Accurate forecasting of, say, wage-rates requires detailed knowledge of:

(i) the wages structure in each major industry,

(ii) the relative bargaining power of labour unions and employers under varying political and economic conditions,

(iii) the timing of applications for wage increases and seasonal wage awards,

(iv) the degree of relationship in each industry between wage-rates (as distinct from earnings) and the movement in retail prices (which may themselves have to be forecast), and

(v) the construction of the index itself—first, its intention; second, its general composition, formula and weighting system; and third, the likely treatment of 'special cases.'

Regarding the last point, it can be most annoying to discover that the only discrepancy between the forecast and actual value of the index has been due to faulty prediction, not of wage changes themselves, but of the treatment which the compilers of the index would be likely to give to particular wage awards meriting special treatment. The problem is most acute for coal and textiles but can also arise where wage awards consist of flat increases rather than percentage changes, where there is a reorganisation of the wages structure, where there is confusion between time- and piece-rates, or where the nature of the data available changes.

A critical study of past movements in the index, and of predictions which might reasonably have been made in the light of information available beforehand, has made it possible to estimate changes in an index of this type with a high degree of accuracy for the oncoming twelve months. But knowledge of index number construction would be a poor substitute for a well-informed analysis of the structure of wage-rates in British industry.

The same sort of problem arises in forecasting long-term cost trends for a particular firm. Once again this means adequate statistics based on information which is both reliable and as detailed as possible, plus a determination not to use statistical elaboration to cloak inadequate material. Statistical elaboration is useless without sound judgment and both are useless without adequate statistical and descriptive data. Given all three, we may expect to do even better next time.

But perhaps the most essential requirement of forecasting, both for management and the statistician, is a sense of humour—it will find full rein in this field.

THE ORIGIN AND DEVELOPMENT OF FACTOR ANALYSIS

DOUGLAS F. VINCENT

The National Institute of Industrial Psychology

Factor analysis was developed largely by psychologists for use in psychology, but it is a statistical method that has been applied to other fields. Mr Vincent's historical exposition should therefore have wide interest.

The Aim of Factor Analysis

The term 'factor analysis' is employed to describe a variety of analytical techniques. The question as to its origin depends upon what techniques are considered to be covered by this term. In my opinion a good definition of factor analysis is: Techniques, developed mainly by psychologists for dealing with their problems, which, by analysing the inter-correlations between sets of measurements, attempt to identify the causes that are operating to produce the variance within each set, and to evaluate the contribution due to each cause.

If a number of children of the same age were given, say, a test of English composition, part of the variance of the scores would be due to general intelligence, part to the differences in verbal ability of the children, part to the imperfections of the test (that is to chance), and part to other causes. Suppose g^2 were the proportion of the total variance due to intelligence, v^2 the proportion due to differences in verbal ability, e^2 the proportion due to chance, and s^2 the proportion due to all other causes. Then if these variances are uncorrelated,

$$g^2 + v^2 + s^2 + e^2 = 1$$

The influences operating to produce the variances g^2, v^2, and s^2 are called 'factors,' and the quantities g, v, and s are known as 'factor loadings.' Any operation for calculating them is a form of factor analysis. Such a calculation would, of course, be quite impossible if the scores of one test only were available, but it can be made from the correlations between the scores of a number of tests given to the same group of children. Actually, no method of factor analysis will determine the s and e loadings. The e loadings can be calculated from the 'reliabilities' of the tests, since e^2 is the difference between the reliability and unity. The reliability of a test is calculated by correlating the scores of parallel versions of it given to the same group of persons at the same time, or from two sets of scores obtained from the same test given to the same group on two occasions with a suitable time interval. The s loadings when they are required are found by difference.

As most methods of factor analysis do not require any advance knowledge of what factors are present, factor analysis is very frequently employed when it is required to find out how many factors actually are present in a group of tests or other measurements, and if possible to identify them, rather than to determine the numerical values of the factor loadings.

If the inter-correlations between the scores of a large group of tests for school subjects were subjected to a factor analysis, it is more than likely that more factors than *g* and *v* would be found. This would mean that in each test the proportion of the total variance s^2, ascribed above to other causes, has been broken up into component parts. Not all of it; in practice s^2 never disappears entirely. A psychologist would endeavour to identify the causes of the other factors; he might ascribe them to influences such as 'the child's reaction towards the teacher and the school' and 'home influence and tradition.' It is a maxim in interpreting the results of a factor analysis that there must be a reason for each factor that appears, and unless a satisfactory explanation can be found for all or nearly all the factors, the factorisation is considered to have failed.

According to the definition given above, all methods of factor analysis are an outgrowth of Galton's concept of the correlation coefficient. Galton showed that if the variances of two sets of measurements can each be regarded as made up of two parts, a part A^2 due to a cause operating in the case of both sets of measurements, and a part X^2 due to one cause in one set of measurements and a different cause in the other set, the correlation coefficient between the two sets of measurements is

$$\sqrt{\left(\frac{A_1^2}{A_1^2 + X_1^2}\right)} \times \sqrt{\left(\frac{A_2^2}{A_2^2 + X_2^2}\right)}$$

that is to say, the geometric mean of two variances. (The term 'variance' came into use long after Galton's time, but it is convenient to use it here as it is the term that is most familiar today.) If a_1^2 and a_2^2 represent the *proportions* of the variances due to the common cause, then

$$a_1^2 = A_1^2/(A_1^2 + X_1^2) = A_1^2/\sigma_1^2$$

and

$$a_2^2 = A_2^2/(A_2^2 + X_2^2) = A_2^2/\sigma_2^2$$

where σ_1^2 and σ_2^2 are the variances of the two sets of measurements. Thus the correlation between the two sets of measurements is

$$\sqrt{(A_1^2 A_2^2/\sigma_1^2\sigma_2^2)} = a_1 a_2$$

That is to say, the correlation coefficient is the product of what are today called the factor loadings. This principle can be extended to the case when there is more than one common cause. If a_1^2 and a_2^2 are the proportions of the total variance of the two sets of measurements due to one common cause, b_1^2 and b_2^2 are the proportions due to

another common cause, and c_1^2 and c_2^2 are the proportions due to a third common cause; then, if the three causes operate independently, the correlation coefficient will be $a_1a_2 + b_1b_2 + c_1c_2$. For two tests similar to that used as an example above the correlation coefficient would be $g_1g_2 + v_1v_2$.

The simplest case of factor analysis is when there is only one common cause and, therefore, only one factor is present. If the inter-correlations of four sets of measurements have been calculated they can be arranged in the form of Table I. Since there is only one common cause, each

TABLE I

—	r_{12}	r_{13}	r_{14}
r_{21}	—	r_{23}	r_{24}
r_{31}	r_{32}	—	r_{34}
r_{41}	r_{42}	r_{43}	—

TABLE II

(m_1^2)	m_1m_2	m_1m_3	m_1m_4
m_2m_1	(m_2^2)	m_2m_3	m_2m_4
m_3m_1	m_3m_2	(m_3^2)	m_3m_4
m_4m_1	m_4m_2	m_4m_3	(m_4^2)

of the correlation coefficients can be expressed as the product of two factor loadings, as in Table II. The quantities in brackets, m_1^2, m_2^2, m_3^2, and m_4^2, which are called the 'communalities,' are not known. If the correlation coefficients were free from errors three sets of measurements would be enough to determine their factor loadings, since

$$m_1 = m_1m_2 \cdot m_1m_3/m_2m_3 = r_{12} \cdot r_{13}/r_{23}, \text{ etc.} \qquad \ldots\ldots(1)$$

Correlation coefficients do, however, contain errors, and the error of a factor loading obtained in this way would be rather large. The problem is how to minimise the errors by averaging all the possible estimates of each factor. It would not involve much computation to work out each estimate separately when only four tests are involved, but when there are a dozen or more tests the amount of work would be very great indeed. A number of devices have been used to obtain an average value for each factor loading with a reasonable amount of work.

If c_1, c_2, c_3, and c_4 are the sums of the columns of Table II and S is the sum of the whole table,

$$\begin{aligned} c_1 &= m_1^2 + m_1m_2 + m_1m_3 + m_1m_4 \\ &= m_1(m_1 + m_2 + m_3 + m_4) \end{aligned}$$

and $$c_2 = m_2(m_1 + m_2 + m_3 + m_4) \text{ etc.}$$

From this it follows that

$$\frac{c_1}{\sqrt{S}} = \frac{c_1}{\sqrt{(c_1 + c_2 + c_3 + c_4)}} = \frac{m_1(m_1 + m_2 + m_3 + m_4)}{\sqrt{(m_1 + m_2 + m_3 + m_4)^2}} = m_1$$

If the communalities were known they could be entered in the blank spaces in a table such as Table I and the averaged estimates of the factor loadings obtained from this table by the formula

$$m_1 = c_1/\sqrt{S} \qquad \ldots\ldots(2)$$

The device most commonly used today is to make a guess at the

communalities and to use this formula. It is not the most accurate method but it is accurate enough in most cases.

The Beginning of Factor Analysis

The first factor analyses ever to be made were all of this simple one-factor type. The reason is historical: they were a direct outgrowth of Galton's conception, which envisaged only the case when only one common cause was influencing the variances of two sets of measurements. Also, Galton's idea fitted in very well with the ideas of the man who began what has since developed into factor analysis.

In his autobiography[1] Spearman tells how, about 1901, he became interested in Galton's work and saw how he could use it to test his own ideas about the nature of intelligence. In 1902 he collected suitable data: test scores and class marks of the pupils at two schools in England. In 1904 he published two papers in the *American Journal of Psychology*; both of them were landmarks. The first[2] was his *credo*, in which he explained and defended his statistical methods. Here for the first time a distinction was made between the statistical treatment appropriate to measurements that can be made accurately, such as the heights and weights and limb measurements of the orthodox statisticians, and the statistical treatment required for measurements that must always contain a large amount of error, such as examination marks, scores at psychological tests, and subjective estimates of temperamental characteristics—'statisticoids' he called them. This paper is the book of Genesis of much that is called psychometrics today. Spearman's second paper[3] was the first publication describing experimental work using the method of dealing with 'statisticoids' that has since developed into factor analysis.

This second paper not only uses the word 'factor' for the first time with its present-day meaning, but it contains a list of factor loadings that Spearman had calculated for eleven measurements. He shows, as an example, a table of the inter-correlations of six sets of measurements arranged as Table I, from which some of his factor loadings were calculated. His first attempt at factor analysis was surprisingly good. His table has been factorised by modern methods, and the six factor loadings obtained agree very well with those in his list.

It is unfortunate that in this 1904 paper Spearman gives no indication at all of the method he used. Apparently he was interested only in obtaining numerical data to support his theory of the nature of intelligence and he did not regard his computational methods as being of any importance. It is certain, however, that in the early days he had devised more than one method of factorising. Among the Spearman documents in the library of the National Institute of Industrial Psychology there is a letter written in 1909 in which he states that there are various formulae which can be used, depending upon the degree of exactness aimed at, but that the results are practically the same in any case.

Only one of these methods has been preserved, and for that we have to thank Sir Cyril Burt. In a paper in 1909[4] he has factorised a table of correlations by using a formula obtained from Spearman, and he has given this formula in a footnote. It is

$$m_s = \frac{a_s}{\sqrt{(2\Sigma - a_s)}} \cdot \sqrt{\frac{n-2}{n-1-na/2\Sigma}} \qquad \ldots\ldots(3)$$

Here m_s is the factor loading of the measurement or test s, a_s is the sum of a column of a table such as Table I (where the communalities are omitted), 2Σ is the sum of all the correlation coefficients in the table; and n is the number of tests or measurements.

This formula is of interest as it is the first factorisation formula to be published. The second term is a correction factor which tends to unity as the number of tests becomes large. The first term of this formula can be used alone without much loss of accuracy. It will be noted that it has a close resemblance to equation (2) and was doubtless obtained by a similar process of reasoning. Actually this early formula is remarkably accurate; in fact it is more accurate than some of the methods most frequently used today. As an illustration of its accuracy an artificial table of the 'correlation coefficients' of six imaginary tests having factor loadings of 8, 7, 6, 5, 4, and 3 has been factorised by this early formula, and by a modern method using guessed communalities. An artificial table is necessary to get 'correlation coefficients' free from error. The factor loadings obtained are as in Table III.

TABLE III

True loadings	0·8	0·7	0·6	0·5	0·4	0·3
Loadings obtained with Spearman's formula (without correction factor)	0·762	0·684	0·596	0·511	0·417	0·318
Loadings obtained with Spearman's formula (with correction factor)	0·798	0·705	0·603	0·508	0·406	0·304
Loadings obtained with a modern method (centroid)	0·756	0·703	0·620	0·532	0·447	0·337

Sir Cyril Burt, again, was responsible for the next factorisation formula to be published. In a memorandum published in 1917[5] he gives in a footnote a formula identical with formula (2) of this article but does not explain how it has been obtained. This footnote suggests that the factor loadings of his tables were obtained not from this formula but from a modification of it. There are several possible modifications of the formula—the first term of Spearman's formula, formula (3) of this article, is one of them. No doubt at this time ideas were rather fluid, and everyone had his own. I know of no positive evidence of when the device of using guessed communalities first came into use, but as formula (2) cannot be used without modification except with guessed communalities it is likely that this device was

in use at this time. The fact that Sir Cyril Burt had quoted this formula suggests that he had been using it, if not for the factor loadings of that memorandum then for some previous work.

No new factorisation formula was published till 1927[6], when Spearman published a particularly ingenious formula which will factorise an error-free single-factor table of inter-correlations with perfect accuracy. It is

$$m_s = \sqrt{(a_s^2 - A_s)} / \sqrt{(2\Sigma - 2a_s)}$$

where A_s is the sum of the squares of the correlation coefficients in column s and the other symbols have the same meaning as in his earlier formula. Probably no one uses it today, as it is rather laborious to work out and the increased accuracy is lost with tables of actual correlation coefficients that are not free from error.

The Beginning of Multiple Factor Analysis

In the early days Spearman and his colleagues were interested in one factor only, that associated with intelligence and which they called the general factor *g*. That other factors could be present in their tables of correlations was known from the beginning. In the work that formed the basis of Spearman's second paper in 1904[3] he had found 'a community purely specific' between Latin translation, Latin grammar, French prose, and French dictation. He even made an attempt to evaluate this 'community.' In the early days such communities, or group factors as they came to be called later, were a nuisance to be avoided while the purpose of factorisation was to provide data for research on *g*.

About 1915 there was a change of outlook and psychologists engaged in factor analysis began to take an interest in group factors for their own sake. In Sir Cyril Burt's memoranda previously mentioned[5] considerable space is devoted to the probability of other factors being present in his measurements and an attempt was made to evaluate them. In 1916 Sir Godfrey Thomson[7] published a paper showing that it was theoretically possible that *g* was the resultant of a complex of group factors. In 1916 the *British Journal of Psychology*[8] published the doctoral thesis of Miss N. Carey, one of Spearman's students. In it she presented a technique for testing a table of correlations for the presence of factors other than *g* and for identifying which tests or measurements have loadings, though not for working out the actual loadings. Her method does, however, contain the germ of the idea that resulted in one of the first techniques of multiple factor analysis, which in its final form became the so-called 'hollow staircase' method.

Dr Carey's idea was to divide the tests or measurements into groups, all the tests in each group being suspected of containing some particular factor in addition to the general factor *g*. Then a table like Table I was made up in which the tests of each group were placed together. If there are three such groups of tests the table can be divided into nine

divisions. The factor loadings to be expected in such a table are as shown in Table IV.

TABLE IV

—	$g_2g_1+a_2a_1$	$g_3g_1+a_3a_1$	g_4g_1	g_5g_1	g_6g_1	g_7g_1	g_8g_1	g_9g_1
$g_1g_2+a_1a_2$	—	$g_3g_2+a_3a_2$	g_4g_2	g_5g_2	g_6g_2	g_7g_2	g_8g_2	g_9g_2
$g_1g_3+a_1a_3$	$g_2g_3+a_2a_3$	—	g_4g_3	g_5g_3	g_6g_3	g_7g_3	g_8g_3	g_9g_3
g_1g_4	g_2g_4	g_3g_4	—	$g_5g_4+b_5b_4$	$g_6g_4+b_6b_4$	g_7g_4	g_8g_4	g_9g_4
g_1g_5	g_2g_5	g_3g_5	$g_4g_5+b_4b_5$	—	$g_6g_5+b_6b_5$	g_7g_5	g_8g_5	g_9g_5
g_1g_6	g_2g_6	g_3g_6	$g_4g_6+b_4b_6$	$g_5g_6+b_5b_6$	—	g_7g_6	g_8g_6	g_9g_6
g_1g_7	g_2g_7	g_3g_7	g_4g_7	g_5g_7	g_6g_7	—	$g_8g_7+c_8c_7$	g_9g_7 c_9c_7
g_1g_8	g_2g_8	g_3g_8	g_4g_8	g_5g_8	g_6g_8	$g_7g_8+c_7c_8$	—	$g_9g_8+c_9c_8$
g_1g_9	g_2g_9	g_3g_9	g_4g_9	g_5g_9	g_6g_9	$g_7g_9+c_7c_9$	$g_8g_9+c_8c_9$	—

Dr Carey suggested that the presence of the group factors *a*, *b*, and *c* in Table IV is indicated by the numerical pattern of the correlation coefficients of the table and by the average of the correlation coefficients in each of the divisions. Methods for calculating the loadings of such group factors came much later. However, it was from a table such as this that Dr W. Stephenson, a colleague and former pupil of Spearman, proved the existence of *v*, the first of the factors appearing in mental tests to be accepted by psychologists as having an importance comparable with Spearman's *g*.

It will be obvious that the loadings of the general factor for all the tests in Table IV could be calculated from those divisions that contain only the general factor. For instance,

$$g_1^2 = g_1g_4 \cdot g_1g_7/g_4g_7 = r_{14}r_{17}/r_{47}$$

and a number of other estimates of g_1 could be made, which could be averaged. Since the correlations in the other divisions are of the form $r_{12} = g_1g_2 + a_1a_2$, if the *g* loadings have been calculated terms like a_1a_2 can be found by difference. These can be assembled into a new table and the loadings of the *a* factor found by a single factor method. This looks simple, but considerable refinement was necessary before a practicable method was devised. The technique known as the 'bi-factor' and the 'hollow staircase' method devised by Dr K. J. Holzinger,[9] which is still in use today, works on this principle.

The next, and up to date the most important, phase in the development of factor analysis began in the early nineteen-thirties and was due mainly to the work of Dr L. L. Thurstone.[10] He not only introduced an entirely new factorising technique but also a new conception

of the nature of factors. His technique, known as the 'centroid' method, is the one most widely used today.

In the first stage a table like Table I is factorised by a single factor method as though it contains one factor only. From the factor loadings all the products, like m_1m_2, m_2m_3, etc. of Table I, are calculated. The original table is then treated as though it contains two factors, that is as though

$$r_{12} = a_1a_2 + b_1b_2 \quad \text{and} \quad r_{23} = a_2a_3 + b_2b_3, \text{ etc.}$$

The terms b_1b_2, b_2b_3, etc., are found by difference and are assembled into a new table. The new table is factorised by a single factor method and the operation repeated, and so on. In this way each of the correlation coefficients is expanded into an infinite convergent series so that

$$r_{12} = a_1a_2 \pm b_1b_2 \pm c_1c_2 \pm d_1d_2 \pm \ldots$$

and similarly for the other correlation coefficients. If, say, three factors are present, the fourth term of the series will fall to zero or, in practice, to a value judged not to be significant. Unlike similar quantities arising in other techniques, the quantities a, b, c, d, etc., need not have any factual interpretation. Indeed, since assumptions have been made that are justifiable only if the original correlation coefficients are regarded as mere numbers, there is no reason why a, b, c, d, etc., should have any meaning, although in some psychological work the first of them is often a rough approximation to the g loading.

If, say, three factors are present, the correlation coefficients are each resolved into three terms, so that

$$r_{12} = a_1a_2 \pm b_1b_2 \pm c_1c_2$$
$$r_{13} = a_1a_3 \pm b_1b_3 \pm c_1c_3$$
$$r_{23} = a_2a_3 \pm b_2b_3 \pm c_2c_3$$
$$\text{and so on.}$$

If there are n tests in the table there are $\frac{1}{2}n(n-1)$ such equations. The whole set of equations, however, is indeterminate; there are an infinite number of solutions. The values for the a's, b's, and c's obtained by Thurstone's method are only one possible solution. The great importance of Dr Thurston's contribution to factor analysis is his device by which any one solution can be transformed into *any* possible solution involving the same number of factors. A meaningless set of factor loadings can be transformed into a set of loadings for which there is an acceptable factual interpretation. Factor loadings are vector quantities, that is to say, they have direction (or a quality analogous to direction) as well as magnitude. Forces are a common example of vector quantities, and Dr Thurstone's original device was a graphical method very similar to the parallelogram of forces. Many more refined graphical and non-graphical methods have been developed from his original idea.

It is Dr Thurstone's device for changing from a meaningless set of factor loadings to a set of loadings that can have a factual interpretation that has led to a new conception of the nature of factors. It has shown that it is always possible that any particular set of tests or measurements can be factorised in more than one way, each set of loadings having an acceptable and even a perfectly obvious factual interpretation. Dr Thurstone's own illustration is one of the best. He took a number of cylinders of varying shape and size and made a number of measurements, such as the length of the greatest diagonal, the area of the ends, and the area of the curved surfaces. He factorised these sets of measurements and showed that there were two quite different sets of factor loadings, each having a perfectly obvious factual interpretation. One set of factor loadings corresponded very closely with *length* and *diameter*, the other with *bulk* and *shape*, shape varying from disc-like to rod-like.

Present-day Views

During the past twenty years there has been a gradual change of outlook upon the nature of factors, which has been due to the work of others besides Dr Thurstone. There is no space to mention everyone here, but Sir Godfrey Thomson's contribution cannot be omitted. He was interested in factor analysis from the very beginning, and in 1916 he published a paper[7] which showed that Spearman's conception of the nature of *g* was not the only one possible. He did not show that Spearman was necessarily wrong, but he broadened the outlook of psychologists using factor analysis and started a new train of ideas which he has continued to develop. The ideas that he has embodied in his Sampling Theory[11] have given those who inclined to Spearman's views much food for thought. Probably his contributions, more than those of anyone, have made psychologists appreciate that factor loadings are statistical conceptions, like means and standard deviations, and are subject to the same limitations. A value for the mean height of a man is of very little use without a definition of the population to which it applies. In the same way the *g* loading, say, of a particular test can be misleading without a definition of the *population of tests* to which it applies. If a non-verbal intelligence test is factorised in a battery of other non-verbal intelligence tests the *g* loading obtained is likely, quite apart from experimental errors, to be slightly different from the loading that would be obtained if it were factorised in a battery containing both verbal and non-verbal tests. If the battery were more varied—if, for instance, it contained tests of attainment in those scholastic subjects known to demand intelligence—the *g* loading obtained would be still more different.

Until about the time of the appearance of Dr Thurstone's work factor analysis had been used only by a small band of enthusiasts. During the past twenty years many psychologists and some statisticians in all parts of the world have become interested in factor analysis and a great many new techniques have appeared. They all depend upon

the principles that have already been outlined in this article, and consist mainly in refinements of methods. Reference to all these contributions to factor analysis is clearly impossible in so short an account as this.

Most treatises on factor anlysis, however, contain some reference to Hotelling's 'principal components,' and this article would not be complete if it did not mention the work of Dr H. Hotelling.[12] But first it should be said that Karl Pearson had to some extent anticipated Hotelling's work thirty years earlier.[13] Pearson used the same mathematical device as Hotelling, but he used it for a different purpose. Pearson's object was to provide an improved technique for prediction, such as making the best estimate of height from arm and leg measurements; he does not appear to have had any thought of analysing his variables. There is no reason to suppose that Dr Hotelling was aware of this previous work.

Principal components have some resemblance to factors but the underlying idea is different. Like factors the loadings of the principal components are calculated from the inter-correlations of the tests, but unlike factors there is no question of choosing an acceptable solution from a number of possible solutions; a resolution into principal components is an exact mathematical operation and the result is unique. To obtain this mathematically unique solution certain assumptions have to be made, which in practice mean that either unities or reliabilities are entered in the table of inter-correlations in place of the communalities. Principal components have the undeniable advantages of a unique solution and the elimination of all guesswork, but they have the disadvantage that about half the loadings are negative and that there is nothing corresponding to the specific factor of factor analysis.

From the mathematical and statistical point of view principal components are far more satisfactory than factors, but from the psychologist's point of view they are of little use; it is only in a few rare instances that they have found an application in psychological work. The obstacles are the inevitable negative loadings and the absence of anything corresponding to a specific factor. Neither factors nor components are of any value to psychologists unless they have a factual interpretation, and it is often quite impossible to attach any meaning to a negative loading. Without the possibility of a specific factor it is necessary to assume that any cause operating to produce a variance in any one of the tests, or other measurements, must also influence the variances of all the others in the battery. In most psychological work this is usually an absurd assumption. If, as an illustration, there is a test of arithmetic in the battery, one cause operating to affect the variance of the scores is the ability to add correctly; it would be absurd to assume that it also affects the scores of any test that does not contain arithmetical questions.

It has been stated that factor analysis is a subject of controversy between the mathematical statisticians and the psychologists. This is

not quite the whole truth. There is only a controversy between those who wish to be controversial. Factor analysis is not a purely statistical technique; there is always a certain amount of guesswork in it which no fair-minded person would attempt to deny; but factor analysis is chiefly a research tool, and there is of necessity a lot of guesswork in all research, from astronomy to nuclear physics. Factor analysis is certainly a very treacherous tool in inexperienced hands. Dr Eysenck[14] in a previous article in this journal has described some of the pitfalls that beset the unwary, but he has also told how leading psychologists, working independently and in very different fields, have obtained remarkably similar results. Undoubtedly there have been many controversies about deductions made from the results of work employing factor analysis, which have focused attention on the weaknesses of the method and have made some people feel very dubious about it. If some mathematical statisticians have suggested that psychologists would be well advised to use the more trustworthy orthodox statistical procedures in place of factor analysis whenever possible, that is not being controversial; it is offering well-meant advice. Unfortunately, however, it is advice that often cannot be taken. There are problems, some of them are mentioned in Dr Eysenck's article, in which factor analysis is the only tool that the psychologists can use.

REFERENCES

[1] Murchison, C. (1930). *History of Psychology in Autobiography*, Vol. I, p. 322. Clark University Press, Worcester, Mass.

[2] Spearman, C. (1904). 'Proof and Measurement of the Association of Two Things,' *Amer. J. Psychol.*, **15**, 72.

[3] Spearman, C. (1904). 'General Intelligence Objectively Determined and Measured,' *Amer. J. Psychol.*, **15**, 202.

[4] Burt, C. (1909). 'Experimental Tests of General Intelligence,' *Brit. J. Psychol.*, **3**, 163.

[5] Burt, C. (1917). *The Distribution and Relations of Educational Abilities*, p. 53. King, London.

[6] Spearman, C. (1927). *The Abilities of Man*, p. 16. Macmillan, London.

[7] Thomson, G. H. (1916). 'A Hierarchy without a General Factor,' *Brit. J. Psychol.*, **8**, 271.

[8] Carey, N. (1916). 'Factors in the Mental Processes of School Children,' *Brit. J. Psychol.*, **8**, 170.

[9] Holzinger, K. J. (1935). *Preliminary Reports on Spearman-Holzinger Unitary Trait Study*. No. 5. *Introduction to Bi-factor Theory*. Chicago University Press.

[10] Thurstone, L. L. (1933). *A Simplified Multiple-factor Method*. Chicago University Press.

[11] Thomson, G. H. (1927). 'A Worked-out Example of the Possible Linkages of Four Correlated Variables on the Sampling Theory,' *Brit. J. Psychol.*, **18**, 68.

[12] Hotelling, H. (1933). 'Analysis of a Complex of Statistical Variables into Principal Components,' *J. Educ. Psychol.*, **24**, 417 and 498.

[13] Pearson, K. (1901). 'On Lines and Planes of Closest Fit to Systems of Points in Space,' *Philos. Mag.*, **2**, 559.

[14] Eysenck, H. J. (1952). 'The Use and Abuse of Factor Analysis,' *J. Appl. Statist.*, **1**, 45.

THE ECONOMICS OF SEQUENTIAL SAMPLING PROCEDURES FOR DEFECTIVES

DAVID G. CHAMPERNOWNE
Nuffield College, Oxford

Inspection is undertaken to reduce 'decision costs' incurred through wrongly accepting or rejecting batches of manufactured articles, and for any given type of sampling scheme these costs decrease as the amount of inspection, and hence the inspection cost, increases. Professor Champernowne deals with the problem of determining optimum sequential sampling schemes which minimise the sum of the decision and inspection costs.

Introduction

During the recent war considerable publicity was given to various methods of quality control designed to reduce waste in industry due on the one hand to unnecessary production and acceptance of goods of too low quality, and on the other to insistence on excessively high standards of quality and the rejection of articles which were satisfactory for their purpose. In particular, attention was drawn to various schemes by which the decision whether to accept or reject a batch of components could be based on inspection of a sample. It was demonstrated that a considerable saving in inspection costs could be effected if, instead of always inspecting the same number of components, the inspection might be discontinued whenever the results of the first few inspections strongly indicated such low or such high quality that the correct decision, rejection or acceptance of the batch, was already fairly obvious. Schemes making use of this principle that the number of components inspected might depend on the results of the inspections so far carried out were known as sequential sampling procedures.

In order to choose the most economical sequential sampling procedure prior knowledge is required under various heads which we shall discuss in the next section. Briefly stated, this prior knowledge is concerned with

(i) the average quality of the batches to be tested and the variation between batches of quality about that average,

(ii) the cost of inspection and its dependence on the amount of inspection undertaken, and

(iii) the cost involved by deciding wrongly to accept or wrongly to reject a batch, and the way this cost depends on the quality of the batch.

Our knowledge under the first heading is bound to be insecure, since we have not yet tested the batches. Many writers, therefore,

preferring not to seek for the most economical sampling procedure, have concentrated on the problem of reducing the risks of wrong decisions to some specified level. The point of view adopted in this article is that by basing one's judgment of the distribution of quality to be expected on the results of tests already carried out on other batches, one eliminates any personal element in such judgment and can expect to choose a more economical scheme than by the orthodox methods, except where the experience with earlier batches turns out to have been wildly misleading. This risk of loss due to a sudden change in the composition of the material is, however, what the scientist must always face in order to avoid the higher penalties that would usually be incurred by ignoring past experience.

The Efficiency of a Sequential Sampling Procedure

In this article we shall give instructions for finding the most economical scheme when the prior knowledge included under the three headings given above is available in a simple form. By the most economical scheme we shall mean that one whose 'expected costs' are the least possible, taking into account both (i) costs of testing and (ii) decision costs, i.e. costs due to wrongly accepting or wrongly rejecting a batch of components. We shall then go on to measure the efficiency of schemes which have been recommended in three practical cases. The measure of the efficiency of a scheme which we shall adopt is obtained by comparing the 'expected costs' of the given scheme with those of the most economical scheme; more exactly, we divide the 'expected costs' of the most efficient scheme by those of the given scheme in order to obtain the measure of efficiency. It will be clear that this measure can range from very small values obtained for very costly schemes to the value of unity obtained for the most economical scheme. The particular schemes which we shall examine will all be found to have high efficiency.

Six Points to consider when Choosing a Sequential Scheme

To simplify the problem let us suppose that a manufacturer must decide whether to accept or reject each of several batches of components each of the same size. Testing is to be carried out one by one on individual components, and the result of any test will be simply 'satisfactory' or 'defective.' Similar tests have been carried out in the past on other batches presumed to be similar to those now to be examined.

Our preliminary discussion suggested that the first task is to estimate from past experience the average quality and variation of quality among batches. In our simple situation the quality of any one batch is given by the number of defectives in relation to the batch-size. The average and variation of quality may be described by the frequency distribution of the number of defectives in batches, provided the batch-size is known. The method of fitting this distribution by means of finding two parameters α and β is explained in the Appendix, and it

depends on finding estimates of the mean and standard deviation of the frequency distribution of the number of defectives. The algebraic expression of the distribution is also given in the Appendix, but all of it that is required for our purposes is the specification of the two parameters α and β. To illustrate the type of frequency distribution obtained we show in Fig. 1 the frequency distributions for batch-size 40, in the cases (i) $\alpha = 8$, $\beta = 0$ and (ii) $\alpha = 12$, $\beta = 1$.

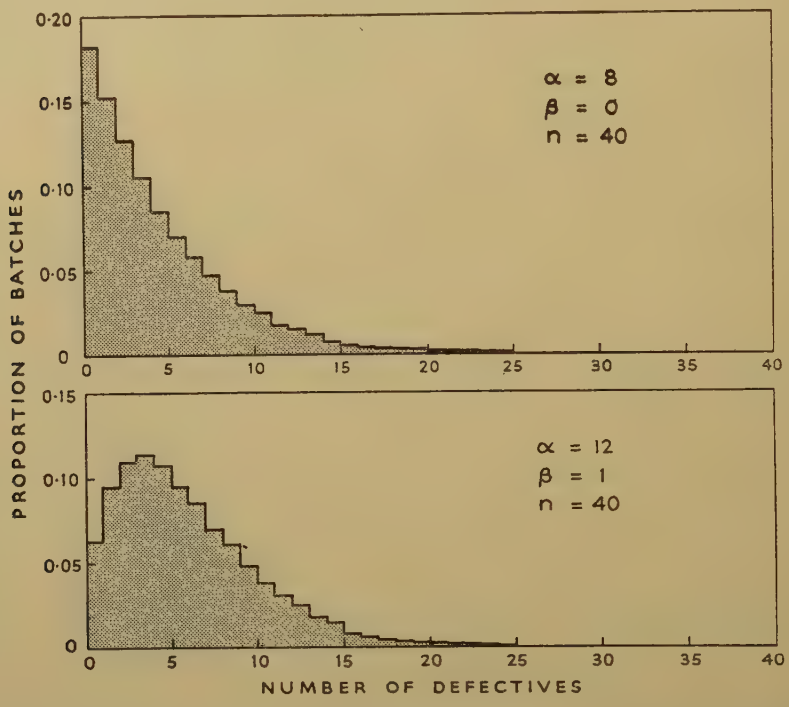


Fig. 1. Frequency distributions of batches classified according to the number of defectives in the batch.

The earlier discussion also indicates that the second task is to estimate how the magnitude of the decision loss involved by a wrong decision depends on the proportion of defective items in the batch. For the purposes of easy exposition we shall assume that the decision costs are of the following simple type. If the proportion defective is in excess of a certain critical proportion p_0, then acceptance involves a cost proportional to that excess, whereas if the proportion defective falls short of p_0, rejection involves a cost proportional by the same rule to the deficiency.

Fig. 2 illustrates the operation of the rule. Here the critical proportion p_0 is taken as 0·1, and the factor b, relating the decision cost to the excess or deficiency of the proportion defective, is taken as 20. We read off from the diagram that if m/n, the proportion defective in the batch, is 0·2, acceptance involves a decision cost of 2; if $m/n = 0{\cdot}05$,

then rejection involves a decision cost of 1. For purposes of finding the most economical scheme we need estimates of the critical proportion p_0 and of the factor b. These must be supplied by the manufacturer after considering what would be the consequences of wrongly accepting a batch with many defectives or rejecting a batch with few defectives.

Again, to make exposition clearer, we shall assume that the cost of testing a number of components in a batch consists of an overhead element C, involved if we test at all, plus a cost of so much, a say, per item tested. The manufacturer requires to know the cost a per additional item tested, but all he need know about C is whether it is sufficiently small for testing to be worth while at all.

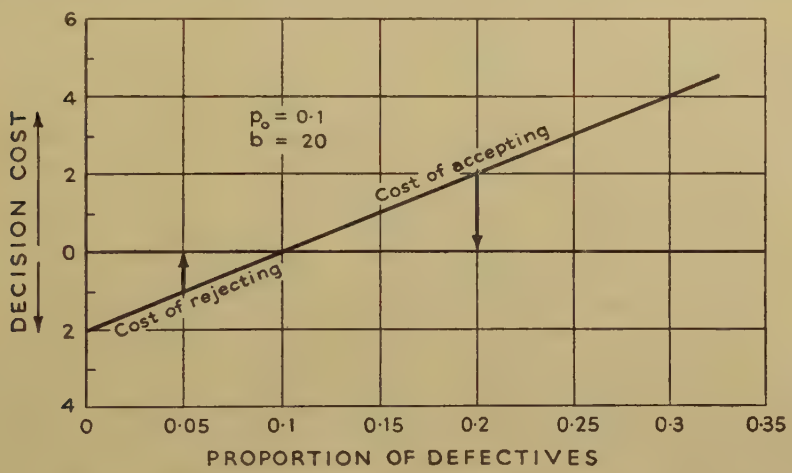


FIG. 2. Graphical representation of the cost of making a wrong decision. Note that when the wrong decision is a rejection, the cost increases as the scale moves downwards from the zero point.

Under the conditions which we are assuming the most economical scheme is always comparatively simple, and its choice can be made a matter of routine once the manufacturer has estimated

(i) the number n of components in each batch,

(ii) the mean and dispersion of the frequency distribution of the numbers defective in the batches, and hence the two constants α and β,

(iii) the critical proportion-defective p_0 and the decision cost factor b, and

(iv) the cost a per additional item tested.

He need not, therefore, trouble to lay down any criteria of simplicity unless he requires something very crude indeed.

The most economical sequential scheme, under the conditions which we are assuming, is always one which can be represented on an acceptance-rejection chart. For the benefit of any readers unfamiliar with these charts one is shown in Fig. 3. On such a chart, the progress of the inspection of a batch is traced by starting from the point o and

advancing one step to the right whenever a component is tested and found satisfactory but one step upwards when a component is tested and is found defective. Thus after three items had been found satisfactory and one defective the point with co-ordinates (3,1) would have been reached. Inspection is to be continued until a step has been taken either across the acceptance boundary, in which case the batch is to be accepted without further testing, or across the rejection boundary, in which case the batch is to be rejected without further testing.

Provided that this type of procedure is sufficiently simple for the manufacturer his only remaining task is to choose the most economical

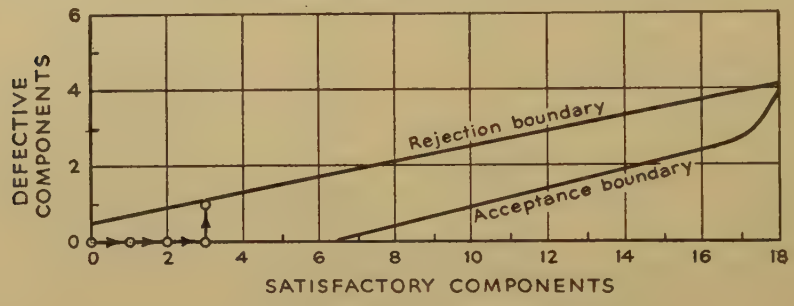


Fig. 3. A typical sequential acceptance-rejection chart.

pair of boundaries. Sufficient evidence for him to make the correct choice is contained in the values of the six parameters we have already discussed at length, n, α, β, p_0, b, and a. Indeed, the form of the boundaries can be found from the four parameters n, p_0, b, and a, and a very simple adjustment in the position of the boundaries on the chart will then be sufficient to allow for α and β.

In the following explanation of the procedure to be followed no proofs will be given of the formulae quoted. These form the subject-matter of articles on the mathematical theory of optimum sequential sampling schemes which are intended for later publication elsewhere. The underlying mathematical theory owes much to the suggestions of Mr A. M. Turing of Manchester University and to the philosophical basis provided by A Wald.[4] Further light on the subject is provided in the works of P. J. Anscombe[1,2] and J. Sittig.[3]

The Form of the Boundaries when the Batches are Large

To determine the form of the most economical boundaries from a knowledge of the four parameters would itself be a very formidable undertaking were it not possible to take advantage of some remarkable properties which these boundaries possess. The first is that their form depends only on two quantities π and θ where, provided the batch-size is large,

$$\pi = p_0 \quad \text{and} \quad \theta = bp_0/a$$

If the batch-size is not large the formulae are more elaborate and they are given in the Appendix.

The second remarkable property of the most economical boundaries is that they are very nearly straight lines. This means in practice that one need only look up five numbers in a table in order to draw both boundaries so accurately that the loss due to any discrepancy between them and the most economical boundaries is negligible. An abridged table and instructions for using it to find the most economical boundaries corresponding to a given pair of values π and θ are given in the Appendix.

To illustrate the straightness of the most economical boundaries, Fig. 4 shows the boundaries obtained for four typical values of π when

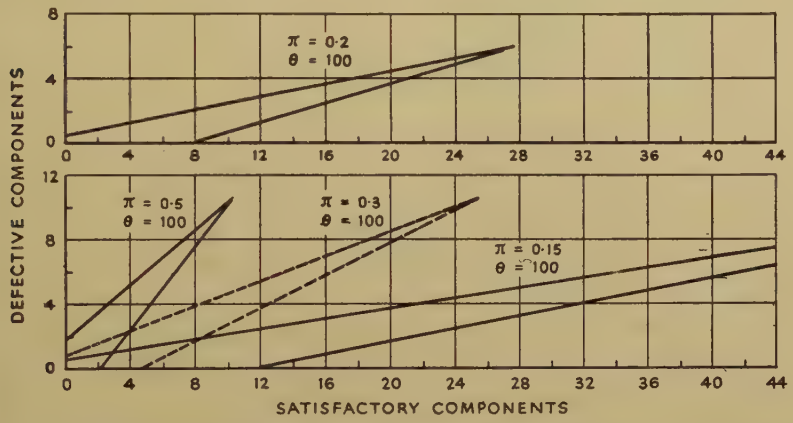


Fig. 4. Some typical sets of most economical acceptance-rejection boundaries.

$\theta = 100$. Table I gives the co-ordinates of the accurately determined acceptance-points and rejection-points for these and some additional values of π and θ; the acceptance-points are such that the acceptance boundary passes just inside (to the left of) them, and the rejection-points are such that the rejection boundary passes just inside (to the right of) them.

It will be noticed that in all these cases the two boundaries meet. This is in striking contrast to the type of boundaries often advocated which consist of a pair of parallel straight lines. The reason that the most economical boundaries always meet is as follows. We are allowing the possibility that the proportion defective in a batch *may* be such that it might as well be accepted as rejected, i.e. either decision would involve negligible cost or no cost. Now, if testing has gone on for a long time without either acceptance or rejection being indicated by a sensible scheme, then this is a strong indication that we are investigating just such a batch; hence it probably will not cost much if we make the wrong decision and it is wasteful to go on testing any more. The calculations leading to the selection of the most economical boundaries take account of this argument expressed in a rigorous form.

The boundaries found from the table for any set of values of n, p_0, b, and a, are those which are appropriate when $\alpha = \beta = 0$. The

TABLE I

Some typical sets of acceptance-points and rejection-points

π	θ	Values of y:	0	1	2	3	4	5	6	7	8
0·5	100	Acceptance, x_y	3	3	4	5	6	7	7		
		Rejection, x_y'	—	—	0	1	2	3	4	6	
0·3	100	Acceptance, x_y	5	7	9	11	13	15	17		
		Rejection, x_y'	—	0	3	5	8	11	13	16	
0·2	100	Acceptance, x_y	8	11	15	18					
		Rejection, x_y'	—	2	7	12	17				
0·15	100	Acceptance, x_y	12	18	23	28	33	38	43	48	
		Rejection, x_y'	—	2	8	15	22	28	35	41	47
0·10	200	Acceptance, x_y	18	27	35	43	50				
		Rejection, x_y'	—	7	17	28	39	50			
0·03	800	Acceptance, x_y	70	100	130	159	188	218	247	278	
		Rejection, x_y'	—	23	61	100	137	174	210	248	285
0·02	800	Acceptance, x_y	96	142	185	229	274				
		Rejection, x_y'	0	50	110	168	223	276			
0·012 5	800	Acceptance, x_y	140	210	284						
		Rejection, x_y'	10	107	204	294					

appropriate allowance for other values of α and β is simply to shift the boundaries to the left by α units and downwards by β units. The same effect can be obtained more simply by starting from the point (α,β) instead of from the origin of the chart at the outset of inspection of each batch. Fig. 5 illustrates the procedure for $n = 1000$, $p_0 = 0{\cdot}3$, $b = 3000$, $a = 3$, $\alpha = 8$, and $\beta = 1$.

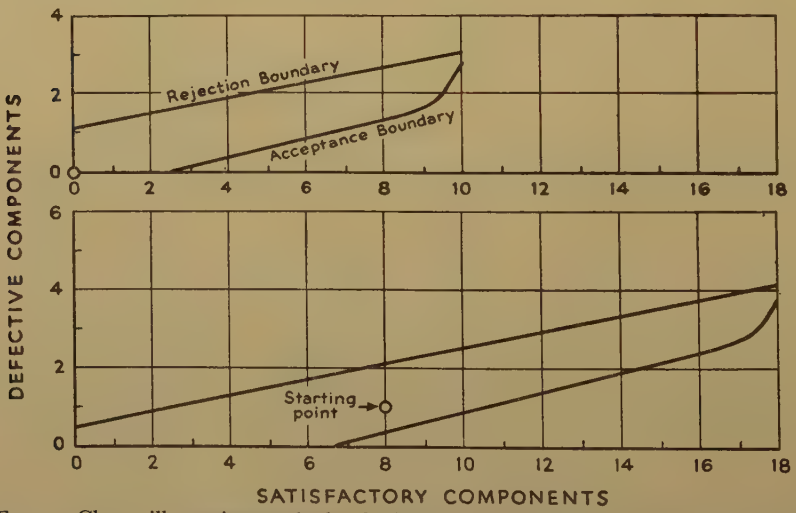


FIG. 5. Charts illustrating methods of allowing for values of α and β not equal to zero ($\alpha = 8$, $\beta = 1$). In the upper chart the boundaries appropriate to $\alpha = \beta = 0$ are moved to the left by α units and downwards by β units. In the lower chart the boundaries shown are those for $\alpha = \beta = 0$, but the starting point for the sequential procedure is (α,β).

The Decision whether to Test at All

If the starting point (α,β) lies outside the acceptance boundary, this means that even if there are no overhead costs of testing every batch should be accepted without testing. The point (α,β) will not normally lie outside the rejection boundary, but if it does, this means that the proportion of defectives is known to be so high in most batches that it is not worth while conducting any tests to find acceptable batches; all batches should then be rejected outright.

But even where (α,β) lies inside the boundaries it may not be worth while to start testing if this involves a large overhead cost C. The question depends on a comparison of the costs of outright acceptance or rejection with those of the most economical of those schemes which involve any inspection. Unfortunately considerations of space forbid further discussion of this question here.

Discussion of Three Examples

In order to illustrate the application of the theory we may now consider three practical examples for which testing procedures have been advocated by Sittig.[3] Fortunately the simplifying assumptions made by him bring these examples within the scope of our methods and he provides estimates of all the quantities which we need to know.

In one important respect Sittig's information is deficient for our purpose. Although in each example he tells us the average proportion of defectives in the batches, he gives no information, beyond a simplifying assumption, about how the proportion varies (from batch to batch) about this average. His assumption is that the frequency function showing this variation is of the form $f(p) = (\alpha + 1)(1 - p)^{\alpha}$, where α is fitted so as to give the correct average. This is equivalent to assuming that $\beta = 0$ in our notation; accordingly in discussing the three examples we shall put $\beta = 0$ without further explanation.

(i) *Toughened Glass.* A simple example discussed by Sittig concerns batches of 200 sheets of toughened glass, in which 10 on the average are defective, used for windows in cars. Here the predominant cost of testing arises from the destruction of the tested sheet, so that taking as our unit the value of one sheet $a = 1$.

If a batch is rejected the glass is re-annealed and re-toughened at a total cost of 100 units; if a batch is accepted, then any defective sheets cause danger to motorists which Sittig values at 10 units per defective sheet. It follows that the critical proportion of defectives is 0·05, at which level either acceptance or rejection will involve a cost of 100 units, so that there is no cost of a wrong decision. At any other level the cost of wrong decision will then be 20 units for every 0·01 by which the proportion defective differs from the critical level of 0·05. Thus the decision-cost function is the line

$$\text{cost} = 2000(p - 0{\cdot}05)$$

so that $b = 2000$ and $p_0 = 0{\cdot}05$.

Since the average number defective is 10 in 200 it follows that $(\alpha + 1)/(\beta + 1) = 190/10$. Using Sittig's assumption that $\beta = 0$, we obtain $\alpha = 18$. The values of the various parameters are accordingly

$$n = 200,\ \alpha = 18,\ \beta = 0,\ a = 1,\ b = 2\,000,\ p_0 = 0{\cdot}05$$

If we adopt the approximate formulae suitable for large n we find

$$\pi = 0{\cdot}05 \text{ and } \theta = bp_0/a = 100$$

With the aid of the more exact formula given in the Appendix these values are adjusted to

$$\pi = 0{\cdot}05 \text{ and } \theta = 110$$

The most economical boundaries can be found by interpolation from the table for $\pi = 0{\cdot}05$ given in the Appendix. If $\alpha = \beta = 0$, the acceptance boundary passes through (28·24,0) and the rejection boundary through (32·23,1) (since 32·23 exceeds 28·24 we need consider no further points). In our example $\alpha = 18$ and $\beta = 0$, so that the boundaries must be shifted by 18 units to the left. They now pass through (10·24,0) (acceptance) and (14·23,1) (rejection). This indicates that the most economical scheme is the following:
Continue testing until either

(i) 11 sheets have been tested and none found defective, in which case the batch should be accepted; or

(ii) a sheet is found defective, in which case the batch should be rejected.

It can be shown that the expected cost of testing and of wrong decision under this scheme is in total 21·71 per batch of 200 sheets. Sittig recommends a slightly different procedure, namely to take 9 sheets and test them all; if any are defective the batch is to be rejected, if none it is to be accepted. Under his scheme the testing cost is 9 and the expected wrong-decision cost is 14·405, giving a total cost of 23·405 per batch of 200 sheets. Comparing this with the expected cost under the most economical scheme, the efficiency of Sittig's scheme is found to be given by $21{\cdot}71/23{\cdot}405 = 0{\cdot}928$. The slight loss of efficiency might possibly be justified on the grounds of the greater simplicity of his scheme, which involves always testing the same number (nine) of sheets in every batch.

(ii) *Radio Potentiometers.* A second example discussed by Sittig concerns batches of 500 radio potentiometers containing on the average 12·5 defective items per batch. Testing costs were 0·05 units per item tested. Sittig's data on the consequences of wrong decisions indicate that a wrong decision involves a cost of 2·7 for every 0·01 by which the batch proportion differs from a critical proportion of 0·092 6. This means that the decision-cost function is $270(p - 0{\cdot}092\,6)$, so that $b = 270$ and $p_0 = 0{\cdot}092\,6$. The testing cost function gives $a = 0{\cdot}05$, and the fact that the average proportion of defectives is 0·025 tells us that $(\alpha + 1)/(\beta + 1) = 39$. Putting $\beta = 0$ we obtain $\alpha = 38$.

If we adopt the approximate formula appropriate to large batches we obtain

$$\pi = p_0 = 0{\cdot}092\,6 \qquad \theta = bp_0/a = 500$$

If we use the more accurate formula of the Appendix which allows for the limited size of the batch, these values are modified to

$$\pi = 0{\cdot}087\,6 \qquad \theta = 511$$

When we find the most economical boundaries to these values of π and θ from tables like those in the Appendix, we find that the point (38,0) lies outside the acceptance boundary. Since $\alpha = 38$ and $\beta = 0$, the most economical scheme is to start from this point (38,0) using these boundaries; this implies that the most economical scheme is to accept without testing. This is indeed the scheme which Sittig selects.

(iii) *Brush-spindles.* The third scheme was concerned with the acceptance or rejection of batches of 500 brush-spindles, containing on the average 11 defective items each. Inspection costs were 0·079 guilders per item and Sittig's discussion of the consequence of wrong decisions shows that the cost of a wrong decision is 19·5 guilders for each 0·01 by which the proportion defective differs from a critical proportion of 0·020 3. This means that the decision-cost function is $1\,950(p - 0{\cdot}020\,3)$, so that $b = 1\,950$ and $p_0 = 0{\cdot}020\,3$. The inspection cost gives $a = 0{\cdot}079$, and the fact that the average proportion of defectives is $\bar{p} = 11/500$ tells us that $(\alpha + 1)/(\beta + 1) = 489/11 = 44{\cdot}45$, and since $\beta = 0$, $\alpha = 43{\cdot}45$.

Using the formulae for π and θ appropriate for large values of n, we should find that

$$\pi = p_0 = 0{\cdot}020\,3 \qquad \theta = bp_0/a = 500$$

If we use the more exact formulae given in the Appendix inserting the batch size $n = 500$, these values are modified to become

$$\pi = 0{\cdot}020\,4 \qquad \theta = 549$$

The most economical boundaries for these values of π and θ and for $\alpha = \beta = 0$, can be found from extended tables similar to those given in the Appendix; since in this example $\alpha = 43{\cdot}4$ and $\beta = 0$, these boundaries must then be shifted to the left by 43·4 units. When this has been done they are found to pass through the points given in Table II. The scheme indicated by these boundaries may be described as follows:

TABLE II

Co-ordinates of points on most economical boundaries (brush-spindles example)

Acceptance boundary ..	(42·8,0)	(86·3,1)	(129·8,2)
Rejection boundary ..	(17·6,1)	(76·6,2)	(>130,3)

Accept the batch directly either

(i) when 43 spindles have passed inspection without one having failed, or

(ii) when 87 spindles have passed inspection with only one failed, or

(iii) when 130 spindles have passed inspection with only two failed.

Reject the batch directly either

(i) if a spindle is failed when less than 18 have been accepted, or

(ii) if a second spindle has been failed when less than 77 have been accepted, or

(iii) if a third spindle has been failed.

It can be shown that the total of decision costs and rejection costs will under this scheme average 6·555 guilders per batch.

The scheme suggested by Sittig is simpler, namely to test a random sample of 31 spindles out of the batch of 500 and to accept the batch if none fail but reject it if any fail. The average decision-cost under this scheme can be found as 5·035 guilders per batch, and the testing cost is 2·449 guilders for each batch, giving a total cost of 7·484. Comparing this with the total cost, 6·555, of the most economical scheme we see that the efficiency of Sittig's scheme is 6·555/7·484 = 0·876. The slight loss of efficiency is the price which has to be paid for the greater simplicity of the scheme.

REFERENCES

[1] Anscombe, P. J. (1951). 'The Cost of Inspection,' *Statistical Method in Industrial Production*, p. 61. Royal Statistical Society, London.

[2] *Idem* (1930). 'Large Sample Theory of Sequential Estimation,' *Biometrika*, **30**, 455.

[3] Sittig, J. (1950). 'The Economic Choice of Sampling System in Acceptance Sampling.' Paper contributed to the Conference of the International Statistical Institute held in New Delhi, India.

[4] Wald, A. (1947). *Sequential Analysis*. Wiley, New York.

APPENDIX

Formulae for the Parameters

When the batch size is large we represent the frequency distribution of p, the proportion defective, by a Pearson Type I curve

$$f(p) = \text{const.}\ (1-p)^{\alpha}p^{\beta} \qquad \ldots\ldots(1)$$

To give any required mean $\bar{p}$ and standard deviation s_p we put

$$\alpha = \bar{p}\left\{\left(\frac{1-\bar{p}}{s_p}\right)^2 + 1\right\} - 2 \qquad \beta = (1-\bar{p})\left\{\left(\frac{\bar{p}}{s_p}\right)^2 + 1\right\} - 2 \qquad \ldots\ldots(2)$$

If β is given but not s_p we obtain any required $\bar{p}$ by putting

$$\alpha = \{(\beta+1)(1-\bar{p})/\bar{p}\} - 1 \qquad \ldots\ldots(3)$$

When the batch size is not large the function (1) may be replaced by

$$g(p) = K\{n(1-p) + a\}^{(a)}\{np + \beta\}^{(\beta)} \qquad \ldots.(4)$$

where (a) and (β) denote factorial powers; equation (2) must then be modified to

$$a = \bar{p}\left\{\frac{(1-\bar{p})^2 - \bar{p}^2/n}{s_p^2 - \bar{p}(1-\bar{p})/n} + 1\right\} - 2$$

$$\beta = (1-\bar{p})\left\{\frac{\bar{p}^2 - (1-\bar{p})^2/n}{s_p^2 - \bar{p}(1-\bar{p})/n} + 1\right\} - 2 \qquad \ldots.(5)$$

When the size of the batch is not large the formulae given in the text for π and θ must be modified as follows:

$$\pi = (p_0 + \bar{p}\gamma)/(1+\gamma) \qquad \theta = b(p_0 + \bar{p}\gamma)/a$$

where $$\gamma = (a + \beta + 2)/n \qquad \ldots.(6)$$

Tables for determining the Most Economical Boundaries

Table A provides values of the five quantities x_0, Δx_0, x_1', x_2', and $\Delta x_2'$ for values of π and θ in a certain range. The most economical acceptance boundary is given with sufficient accuracy by the equation

$$x = x_0 + y\Delta x_0 \qquad \ldots.(7)$$

and the most economical rejection boundary is given almost exactly by a curve going through the points

$$(x_1',1),\ (x_2',2),\ \text{and}\ (x_r',r),\ r = 3,\ 4,\ 5,\ \text{etc.}$$

where $$x_r' = x_{r-1}' + \Delta x_2' - (r-3)/100p_0 \qquad \ldots.(8)$$

Although the figures are given in the table to one or two places of decimals they are based on approximate methods of calculation, including some interpolation. As the number of computations was large it is not certain that all arithmetical slips have been corrected, but reasonable care has been taken.

TABLE A

Values of x_0, Δx_0, x_1', x_2', and $\Delta x_2'$ for determining the most economical boundaries of an acceptance-rejection chart

θ	π / $1/100\pi$	0·01 / 1·00	0·02 / 0·50	0·03 / 0·33	0·04 / 0·25	0·05 / 0·20	0·06 / 0·17	0·07 / 0·14	0·08 / 0·12	0·09 / 0·11	0·10 / 0·10
50	x_0	n.a.	47·3	n.a.	28·2	23·7	20·2	17·6	15·5	13·9	12·5
	Δx_0	...	...	...	...	...	...	...	...	...	...
	x_1'	...	...	...	...	...	29·7	24·7	20·7	17·5	15·1
	x_2'	...	...	...	...	...	...	...	...	...	...
	$\Delta x_2'$	...	...	...	...	...	...	...	...	...	...
100	x_0	97·3	59·1	43·0	33·8	27·8	23·6	20·3	17·8	16·2	14·7
	Δx_0	...	...	...	...	...	...	...	10·6	9·0	8·0
	x_1'	...	...	62·4	44·1	33·0	25·6	20·4	16·6	13·7	11·6
	x_2'	...	...	...	...	...	...	...	29·9	26·0	22·5
	$\Delta x_2'$	...	...	...	...	...	...	...	...	...	...
200	x_0	120·9	70·4	49·9	38·5	32·5	27·9	24·4	21·7	19·5	18·0
	Δx_0	...	...	...	22·2	17·00	14·08	12·00	10·42	9·20	8·02
	x_1'	...	90·9	53·7	35·6	25·6	19·0	15·0	11·7	9·2	7·3
	x_2'	...	...	...	62·8	48·4	37·9	31·1	25·7	21·6	18·3
	$\Delta x_2'$	...	...	...	...	22·9	18·2	15·1	13·18	11·60	10·34
300	x_0	134·5	76·3	54·5	43·4	36·0	30·8	27·2	24·2	21·8	19·8
	Δx_0	...	...	28·9	21·62	17·24	14·30	11·94	10·40	9·20	8·18
	x_1'	193	81·9	45·8	29·7	20·8	14·8	10·7	8·2	6·3	4·9
	x_2'	...	...	85·0	58·5	43·7	33·7	26·8	21·7	17·9	14·9
	$\Delta x_2'$	...	...	...	27·9	22·7	18·01	15·30	13·26	11·64	10·41
400	x_0	143·6	80·1	58·8	46·4	39·1	33·3	29·1	25·8	23·3	21·3
	Δx_0	...	45·2	29·16	21·84	17·04	14·20	12·16	10·52	9·28	8·28
	x_1'	185	73·6	40·5	25·2	16·2	11·4	8·2	6·0	4·4	3·2
	x_2'	...	129	79·2	54·1	38·2	29·7	23·2	18·5	14·9	12·2
	$\Delta x_2'$	...	...	37·5	27·60	21·88	18·00	15·34	13·30	11·71	10·44

θ	π / $1/100\pi$	0·11 / 0·09	0·12 / 0·08	0·13 / 0·08	0·14 / 0·07	0·15 / 0·07	0·16 / 0·06	0·17 / 0·06	0·18 / 0·06	0·19 / 0·05	0·20 / 0·05
50	x_0	11·4	10·4	9·6	8·8	8·1	7·4	6·9	6·6	6·2	5·9
	Δx_0	...	...	...	...	5·2	4·9	4·5	4·1	3·8	3·5
	x_1'	13·2	11·5	10·1	8·96	7·95	7·1	6·4	5·7	5·1	4·6
	x_2'	...	...	...	...	14·5	13·2	12·1	11·2	10·4	9·6
	$\Delta x_2'$	...	...	...	...	...	...	...	...	...	...
100	x_0	13·5	12·5	11·5	10·7	10·0	9·4	8·8	8·2	7·8	7·5
	Δx_0	7·21	6·54	5·98	5·50	5·08	4·71	4·39	4·10	3·84	3·51
	x_1'	9·8	8·1	7·4	6·4	5·5	4·8	4·1	3·5	3·0	2·6
	x_2'	19·6	17·1	15·6	13·9	12·5	11·2	10·2	9·2	8·3	7·6
	$\Delta x_2'$	9·8	8·7	7·8	7·0	6·4	5·96	5·55	5·19	4·86	4·57
200	x_0	16·4	15·2	14·0	13·0	12·2	11·4	10·8	10·2	9·6	9·1
	Δx_0	7·26	6·60	6·04	5·58	5·12	4·76	4·42	4·14	3·88	3·64
	x_1'	6·0	4·9	4·1	3·4	2·8	2·3	1·8	1·5	1·2	0·9
	x_2'	15·7	13·6	11·8	10·4	9·1	8·0	7·1	6·3	5·5	4·9
	$\Delta x_2'$	9·31	8·45	7·72	7·10	6·56	6·09	5·73	5·30	4·97	4·67
300	x_0	18·1	16·7	15·5	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
	Δx_0	7·37	6·71	6·14	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
	x_1'	3·9	3·0	2·3	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
	x_2'	12·6	10·7	9·1	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
	$\Delta x_2'$	9·36	8·51	7·77	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.

... not required n.a. not available

LETTER TO THE EDITOR

Estimating Individual Expenditure from Family Totals

In his article 'Estimating Individual Expenditure from Family Totals,' published in Vol. I, No. 3, of *Applied Statistics*, Mr W. F. F. Kemsley gives estimates of what purports to be the subdivision of total family expenditure between the different members of the family. His aim, in his own words, is to estimate the 'expenditure on each individual' in the family. The only data he uses are (i) figures of total family expenditure and (ii) the numbers in the family, classified in broad groups. His estimates are obtained by the simple application of least squares regression of (i) against (ii). The data are actually confined to expenditure on food, but the method, as he says, could just as easily be applied to other items, or to total expenditure on all items.

The method leads to generalised estimates of the differences in the total expenditure of families of different composition. For a child under 6, for instance, the method results in an estimate of the average increase in total family expenditure which accompanies an increase by one in the number of children under the age of 6; Mr Kemsley does not distinguish between first, second, or later children in a given age group. There is, however, no reason to suppose that this is the same as, and it is misleading to call it an estimate of, the amount spent on the child. It is likely, in fact, to differ substantially from the amount spent on the child, for reasons mentioned below.

Income and family size are correlated. Larger families may contain, for instance, more earners than smaller families. Mr Kemsley's results simply illustrate the fact that the larger families generally have more money to spend. His estimates show, in fact, how much more, on the average, one type of family spends than another. The method is not capable of telling us how the average expenditure of either type of family is apportioned over the different members of the family. If the addition of another member to the family were associated with a reduction in total expenditure—and in 1938 working-class families with three children spent less than those with one or two children—the expenditure on the additional member, estimated by his method, would be negative. His estimates must reflect, in short, all the factors that are associated with differences in family composition.

Mr Kemsley has not included income, or any measure of the standard of living, in his regression equations. And yet it is necessary to have some yardstick for comparing families of different composition before we can begin to estimate the expenditure on individual members of the family. The estimates I published in the *Journal of the Royal Statistical Society* of 1949, to which Mr Kemsley refers, were based on comparisons of the expenditure of families which were estimated, by the methods there described, to have the same standard of living. A set of regression equations was obtained simultaneously from the expenditure figures, covering the whole range of incomes, of all families which were broadly of the same type, except for having different numbers of children. From the regression equations for the different sized families it was possible to derive equations expressing the

estimated expenditure on the first or second child as a function of the standard of living. The estimates which were quoted related to the average standard of living of all the families in question. The method was thus rather different from what Mr Kemsley calls the subtraction method, which he incorrectly attributes to me and which simply assumes, analogously to his own method, that the difference between the average expenditure of two groups of families of a given type, one having one more child than the other, is the expenditure on the child. Although my estimates related to households of a given type, viz. working-class industrial households containing two adults—one man and one woman, and although it might be possible to generalise the method to include other types of household at the same time, it is doubtful whether the sampling errors of the estimates for this particular type of household would be much reduced.

The standard of living is also correlated with the size of the family. As the size of the family increases, average expenditure does not generally increase *pari passu* with the extra needs of the family, and so the average standard of living tends to decline. When a new member of the family arrives the expenditure on the previous members of the family is liable to be reduced. Mr Kemsley's estimates do not allow for this. As estimates of the expenditure on individuals they are therefore *biased*. They relate, in fact, not to the expenditure on individuals but to the total expenditure of families of different composition. This is a different, much simpler, and to my mind much less interesting problem than that of allocating total expenditure between the different members of a family. The fact that the sampling errors are smaller is cold comfort.

13th January 1953 J. L. NICHOLSON

[Mr Kemsley replies as follows:

'Most of the points that Mr J. L. Nicholson is making were briefly referred to in the introduction to my paper. For the rest, Mr Nicholson is concerned with a different issue, namely the desirability of incorporating in the analysis some measure of the standard of living. This is part of a more general econometric problem which I purposely excluded from the scope of my article, as is clear from the introduction.

Naturally it is to be expected that different results will be obtained if allowance is made for the standard of living. It must be remembered that the amount by which estimates of individual expenditure differ will depend on the item under study, its income elasticity, and the method by which the standard of living is assessed. Unfortunately there seems to be no general agreement as to how the standard of living should be measured, and different methods have been adopted in studies made during the last decade. What is now required is a comprehensive investigation in which the various methods that have been proposed as measures of the standard of living shall be applied to the same data and the results compared. Such comparisons as are available do not appear to have provided any firm conclusions.']

NOTES AND COMMENTS

Too much introspection is perhaps as bad for a journal as for a human being, but readers will be interested to know how *Applied Statistics* is progressing while it is still very young. The journal was established to appeal to readers ranging from the interested layman to the expert statistician, and to people interested in all the main fields of application of statistics. It was also intended to provide a medium for contributors who have experiences and ideas to communicate which, although worthy of propagation, are not substantial enough for the more learned journals. The signs are that *Applied Statistics* is succeeding in these objectives, although there is of course room for improvement and development.

The subscribers to the first volume number 1 469, and of these nearly one-half are from overseas. Subscribers were invited, when renewing their subscriptions, to disclose their fields of interest. Only 91 did so and the sample is probably biased; but all the fields of interest enumerated in the returns were well represented and it is clear that *Applied Statistics* must continue to serve the full range of readers originally addressed.

When stating their fields of interest subscribers were also invited to indicate the types of article they favour. The specified types and the numbers of votes favouring them are as follows: general surveys of fields of application of statistics (38); practical examples of the application of statistical methods (90); descriptions of statistical organisations (23); expositions of well-established statistical methods (47); expositions of new developments in statistical methods (75); detailed descriptions of computing methods and of statistical points that arise in practical application (68). In view of the nature of the sample it is reasonable to infer, and to infer only, that *Applied Statistics* should continue to print articles of all the types mentioned. Even descriptions of statistical organisations, for which enthusiasm is apparently least, should have a place. Intending contributors should note the interests of subscribers, for the articles printed can only be selected from those submitted.

The demand for practical examples of the application of statistical methods is almost unanimous, and it was envisaged at the start that articles of this kind would constitute an important part of *Applied Statistics*. Contributions showing statistics in action, especially in industry, are very welcome; and if they are short and, from a statistical point of view, elementary, so much the better. There is clearly a strong interest among readers in expositions of statistical methods, even of well-established methods, and we hope, with the co-operation of contributors, to serve this interest more adequately in the future than in the past. Care will be necessary in presenting such expositions, especially of established methods, for we do not think that it is a function of *Applied Statistics* merely to reproduce, in effect, slices of a statistical textbook. An acceptable exposition will have, among others, a quality that we can only describe as 'freshness.' The exposition will be illustrated by examples of practical interest and will show why one method is used in preference to an alternative; it will bring out points, especially such as arise in

practice, that are inadequately covered in orthodox textbooks; it will expound methods for novel applications; and so on. We hope that more contributors will try their hands at the difficult but rewarding exercise of exposition.

Articles continue to be submitted in such numbers that we confidently expect to maintain such interest and quality as has already been achieved. Moreover, intending authors can take encouragement from the fact that it has so far been possible to publish all suitable articles with a delay that is not excessive for a journal issued only three times a year.

It is not the aim of *Applied Statistics* to compete with other journals for articles, especially with the *Journal* of the Society, although, in attempting to give readers a balanced and complete diet, we have doubtless published some that would have been also suitable for publication elsewhere. The aim is to give *Applied Statistics* a distinctive character and to fill what would otherwise be a gap in the literature.

★ ★ ★

The Council of the Royal Statistical Society offers the Frances Wood Memorial Prize of £40 for competition in 1954. As the object of the prize is to encourage young investigators, competitors must be under the age of 35 on 1st October 1954. It is hoped that this offer of a prize will appeal specially to the younger Fellows of the Society, but it is also open to those who are not Fellows and in particular to members of the Research, Industrial Applications, and Study Sections of the Society and of their local groups.

The prize is offered for the best investigation, on statistical lines, of any problem, *chosen by the competitor*, which bears directly or indirectly upon economic or social conditions. This phrase may be interpreted widely and might, for example, concern housing, distribution, food production, health or nutrition, conditions of work, output, wages or expenditure, education, or old age. If there exists any doubt about the suitability of a subject, enquiry should be made of the honorary secretaries of the Society, who will seek the views of the Council and advise accordingly.

It is suggested that the investigation should be made from material obtained by original observation or analysis carried out in the course of the competitor's ordinary work or special research. The treatment must be mainly statistical and, where material can with advantage be presented in the form of tables and diagrams, such methods should be freely used. The subjects chosen by previous successful candidates are listed below.

In general the Council does not expect private investigators to produce reports on the scale of those compiled by official bodies employing whole-time investigators but accepts for the competition small-scale pieces of research showing original work and personal effort seriously undertaken. Theses submitted or intended to be submitted as university exercises and for degrees and diplomas are admissible. The essay should be in the region of 5 000 to 10 000 words long.

Entries must be printed, or typed on foolscap or quarto paper, but diagrams may be hand-drawn. Entries must be sent to the Honorary Secretaries of the Royal Statistical Society, 4 Portugal Street, London, W.C.2, not later than 30th September 1954.

Previous awards have been:

1921. Winifred Mackenzie: Changes in the Standard of Living of the Working Classes in the United Kingdom between 1860 and 1914

1923. E. G. M. Haynes: Human Power in the English Pottery Industry

1925. Christopher T. Brunner: An Inquiry into Local Variations in the Birth-rate

1928. Colin G. Clark: A Graphical Analysis of the Unemployment Position, 1920–28

1932. R. W. B. Clarke: Production, Output per Head, Prices and Costs in the Iron and Steel Industry, 1924–31

1936. C. V. Dawe: An Economic Interpretation of the Agricultural Statistics relating to the Bristol Advisory Province

1938. R. Huws Jones: Physical Indices and Clinical Assessments of the Nutrition of Schoolchildren

1940. H. W. Singer: The Process of Unemployment, statistically analysed

1946. J. Marley: A Statistical Survey of Certain Aspects of the Beef Producing, Dairying, and Cattle Rearing Industries in Great Britain between 1939 and 1947

1946. T. Barna: Britain's Manpower Problem in the Next Five Years

1949. J. L. Nicholson: Variations in Working Class Family Expenditure

1949. L. T. Wilkins: The Prevalence of Deafness in England, Scotland, and Wales

★ ★ ★

After discussions extending over about four years, the International Statistical Institute has appointed a Committee on Statistics in Industry and Technology. The tasks of the Committee are: (i) to advise the ISI in all matters related to industry; (ii) to provide industrial contributions to the biennial sessions of the ISI; (iii) to organise and supervise an Annotated Bibliography; and (iv) in general to encourage and promote the application of statistical methods in industry. It is envisaged that (iv) will be furthered by facilitating personal contacts and an exchange of experience between statisticians in different countries; by encouraging the teaching of industrial statistics in underdeveloped countries; by the co-ordination of statistical research in different countries; by standardising symbols and terminology; and by promoting the application of statistical methods in the standardisation of industrial techniques and products. At the outset the committee consists of President: Dr W. A. Shewhart (USA); Chairman: Mr L. H. C. Tippett (UK); Secretary: Dr H. C. Hamaker (Netherlands); Members: Dr W. E. Deming (USA), Professor Dr A. Hald (Denmark), Mr Koyanagi (Japan) and Dr (Miss) S. P. Vaswani (India).

Papers on applications of statistics in industry now appear in so many journals that it is difficult to keep abreast of developments, and a regularly published international annotated bibliography should provide a valuable service to industry.

Plans are taking shape for the next conference of the ISI to be held in Rome during the week commencing 7th September 1953. Among other items there will be sessions on Statistics and Standardisation and Education and Training for Statistics in Industry, and papers are being invited from countries that are active in the various fields. Information may be obtained from the Director of the ISI at 2 Oostduinlaan, The Hague, Netherlands.

INDUSTRIAL APPLICATIONS SECTION OF THE ROYAL STATISTICAL SOCIETY

List and Reports of Meetings

BIRMINGHAM AND DISTRICT GROUP

7th January 1953. Everyday Uses of Probability: M. J. Moroney

28th January 1953. Sources of Variation in a Transformer Sheet Mill: A. D. Grace

25th February 1953. Elimination of Trends: O. L. Davies

Dr Davies outlined various methods of counteracting the effects of trends in planned experiments and dealt in some detail with procedures which involve fitting curves to account for time trends. The discussion which followed mostly concerned the justification and the danger of fitting polynomials.

GLASGOW GROUP

20th January 1953. Statistics in the Explosives Industry: A Symposium

12th February 1953. Statistical Methods in the Hosiery Industry: A. Glaskin

LEICESTER GROUP

29th January 1953. Short-cut Analysis of Variance Based on Range: H. O. Hartley

26th February 1953. The Use of Statistics in Consumer Research: P. Harper

LONDON GROUP

6th February 1953. Selling Statistics to Industry: M. J. Moroney

Mr Moroney drew a parallel with the propagation of sales technique for any commercial product. The four steps regarded as essential are (i) the stimulation of interest, (ii) the dissemination of information, (iii) the focusing of this information on to the customer's needs, and (iv) a statement showing the gain to be derived. At present facilities for the training of statisticians are, in most areas, inadequate. It is therefore necessary for the statistician himself to undertake the 'selling' of his subject.

MERSEYSIDE GROUP

22nd January 1953. Statistics in Glass Manufacture: S. P. L. Truesdale

Mr Truesdale spoke chiefly about methods for controlling the quality of the raw materials. A chart in the usual quality control form of amount of matter insoluble in hydrochloric acid had been in use for a number of years. Over the years 1938–43 it revealed the steady upward trend that occurred during the war years, and also several more abrupt changes which could be associated with changes in the grade of workpeople employed as the war progressed.

A second problem concerned the method adopted for sampling the sand. A week's supply is stored in a single 2000-ton container after washing and grading. This manner of storing caused the sand to be stratified, the quality usually deteriorating with depth. The usual method of taking a sample involved mixing the material taken from several points and thus gave no information about variability. It was also found that the 'sand-getter' was able to arrange the quantities he took from each level so as to get the required iron content in the sample. The method now adopted gave more nearly random samples. This, combined with a more rapid spectrographic method of analysis, now gave more information about the material.

24th February 1953. A Study of Variation in Concrete Strength: B. E. Roberts

NORTH-EASTERN GROUP

21st January 1953. The Adequacy of Underground Communications Systems: B. H. P. Rivett

18th February 1953. Statistical Control as an Aid to Management in Industry: A. W. Swan

SHEFFIELD GROUP

29th January 1953. The Correlation of Time Series, with special reference to Routine Works Data: G. H. Jowett

In carrying out a regression analysis between a measure of a 'cause' and a measure of an 'effect' it is usual to calculate the least-squares regression of y on x, together with the standard error of the slope, the latter serving as a basis for determining limits of the true regression slope. If, however, the variables are time series, this standard error is the usual figure (determined as for random series) multiplied by a factor depending on the autocorrelations or mean semi-squared differences of the two series. This factor may be quite large (e.g. ten or more) in series which do not exhibit stable variation.

In practice the factors are difficult to assess from the series themselves since these, in the typical industrial investigation, are rarely long enough. The factors may be impossible to assess at all accurately when slow oscillating trends are present. It is therefore usually necessary to carry out some form of trend removal on the series before fitting a regression. The simplest technique, which is efficient when the effect variable shows more autocorrelation than the cause variable, is to fit and remove the trends with a moving average of three. The standard error estimate, as usually calculated, then needs to be multiplied by about 1·3.

26th February 1953. Some Statistical Problems in a Game of Patience: S. Peiser

SOUTH WALES GROUP

29th January 1953. Quality Control into Operational Research: A. W. Swan

26th February 1953. Queues, and some Interference Problems in Industry: W. D. Ewan

STUDY SECTION OF THE ROYAL STATISTICAL SOCIETY

Reports of Meetings

LONDON

8th October 1952. The Accountant and Statistician in Business: F. Sewell Bray and C. Nuttall

A number of opposing points of view were ventilated at this meeting and many unsolved problems were described. The meeting succeeded in clarifying some of the problems of liaison between the two professions, and Professor Bray indicated a number of areas in accounting practice where the statistician might be able to provide valuable tools. It was clear that much ground could, with profit, be explored in the border country between accepted accountancy methods and statistics.

12th November 1952. The One per cent. Sample Census: E. Grebenik and J. Mason

Mr Grebenik reviewed the first issues of the one per cent. sample census from the viewpoint of a demographer. He welcomed the information which would be available for work in the field of fertility studies and indicated a number of the problems which arose from peculiarities in the definitions. For example, the classification of widows (since women live longer than men) led to some results which looked quite peculiar on their face value, such as clergymen appearing to have on the average more than one wife each!

Mr Mason pointed out the uses which the businessman could make of the census data, but was particularly critical of the social class definitions and the method of analysis. In the discussion the Chief Statistician of the Registrar-General's Office clarified a number of points raised by the speakers and from the floor. It was evident, however, that a number of errors had avoided detection and remedy. In particular the numbers of working women were suspected because the definition of 'wholly or mainly' engaged on household duties was apparently applied differently by different enumerators. Official comments on differences between the census and other officially published data have since appeared in the *Ministry of Labour Gazette*.

10th December 1952. Statistics of the Post and Mail: Sam Wood

Mr Wood commenced his paper by outlining the system by which the Post Office handles its mail. These were 'statistics' in the sense generally accepted among the lay public—figures of most impressive magnitude. He told about the methods used for maintaining controls on service mechanisms and methods of counting the traffic. After this he turned to a discussion of the possibilities and difficulties in instituting sampling schemes. It appeared that no system of marking of the envelopes for sampling purposes was politically possible, since the public immediately question any unusual marks appearing on their correspondence. Several ingenious methods had

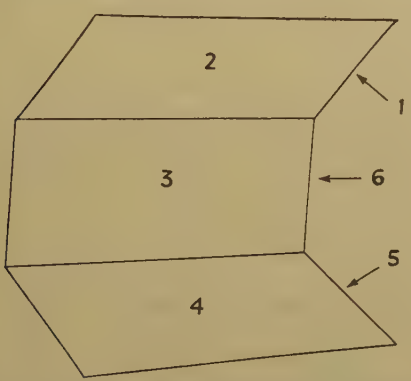
been put forward to overcome this difficulty. One was the use of radioactive spotting of the sampled mail and the use of Geiger counters; this had to be rejected since there was a remote chance of an accumulation of a sufficient quantity of radioactive material to cause danger from the concentration.

The discussion turned its attention to ways and means whereby sampling schemes might be effected and to problems of avoiding bias. Mr Wood outlined a scheme which was being put through its trials and its technical difficulties were examined in detail. If the trials are successful he is hoping to be given permission to report the methods later. It is possible only to state at the moment that the system involves the use of the newer fluorescent paints and infra-red beams.

14th January 1953. The Design of Schedules: D. L. Lamberth

Schedules were described as 'forms with many functions,' and the speaker demonstrated the large variety of purposes which could be served by a well-designed form. The requirements of all who would handle the forms, from its origin to its use by the punch operators and editors, and the time involved, had to be considered. Usually the interviewers' time or the time of the subject receiving the form could be measured in hours; the coders' or analysts' time could be measured in minutes; and the punch operators' time in seconds.

Mr Lamberth showed how savings of time were possible by arranging turning of the forms. For a foolscap-page schedule (a brief sheet once folded) the least accessible page was sheet 2; for a six-foolscap-page schedule (a sheet 23 in. by 13 in. with two parallel folds) the least accessible spaces were those shown as pages 3 or 4 in the diagram. Tumbler printing was used in these layouts.



Page 4 folds to 3; 2 folds to 5.

Pages 1, 2, 5, and 6 can be treated as a four-page schedule.

The speaker also showed how code numberings could be used to save time and increase flexibility of the analyses. He gave an example where the codes in a battery item were in the form of a Latin square, giving a saving of 16 columns in 24 of the punched card. Another device was the preprinting of coding numbers in the coding spaces instead of leaving a blank box for the insertion of the code. Coding was done by circling the appropriate number. This reduced the time of punch operators, and errors of reading. The layout of the numbers in the box was the same as that of the punch operator's keyboard.

STUDY CIRCLE ON MEDICAL STATISTICS: ROYAL STATISTICAL SOCIETY

THIS Study Circle has completed its first full year of life and there is clear evidence that it is fulfilling a real need. Membership is not confined to people from the London area, and statisticians from Edinburgh, Manchester, and Oxford are among the regular attenders. Meetings are held, by kind permission of the Dean, in the Medical School of the Westminster Hospital and normally take place on the last Tuesday of the month. The following are details of meetings held during 1952–3.

30th September 1952. The Validity of Medical Records: M. Russell

Miss Russell argued that in medicine, even more than in other branches of science, it is necessary to examine with particular attention the material on which statistical deductions are based.

28th October 1952. The Spread of Influenza within Households: Ian Sutherland

Dr Sutherland described a survey recently made in a general practice during an epidemic and showed how difficult it was to prove, by the standard methods of statistical epidemiology alone, that influenza was in fact an infectious disease. He suggested that this might be due to the shortness of the incubation period.

25th November 1952. Medical Records in General Practice—Coding and Indexing the Histories and the Work: A. Garvie

Dr Garvie described in detail the very ingenious methods he has evolved over many years and stimulated a useful discussion.

27th January 1953. Sickness Absence Statistics: F. J. Lloyd and K. G. Murden

Mr Murden dealt with the methods used in his organisation, the London Transport Executive, for transferring the basic material from the administrative records to punched cards. Mr Lloyd discussed some of the preliminary results.

24th February 1953. The Nature and Sources of Medical Statistics: B. Benjamin

31st March 1953. Mr J. H. Austin, statistician to the Oxford Regional Hospital Board, read a paper on the work of his department.

28th April 1953. Having a First Baby: Angus Thomson

Dr Thomson described some results of research carried out by the Department of Midwifery in the University of Aberdeen in association with the MRC Social Medicine Research Unit.

26th May 1953. Some Aspects of Mortality in Scotland: Lilli Stein

Another of the Study Circle's activities has been the setting up of small groups to study and report on the following subjects: (i) the measurement of morbidity, particularly in relation to the needs of public health authorities; (ii) the assessment of the need for and the efficiency of clinical services; (iii) the statistical validity of diagnostic procedures; (iv) disease classification; and (v) punched card techniques and their applications.

QUALITY CONTROL IN THE USA

EDWARD D. VAN REST

Statistical Advisory Unit, Ministry of Supply

Quality control is usually regarded in Great Britain as a statistical subject, but Mr van Rest found in the USA a tendency to regard it as a branch of management concerned with the control of quality, control involving not only manufacture to a specification but also a satisfaction of the consumer's needs.

I WENT to the USA as a member* of a specialist team brought together by the Anglo-American Council on Productivity (now British Productivity Council) to study inspection methods and organisation in the light engineering industry.[1] This article describes present practice in this field in the USA as we saw it and discusses the changes in outlook that have occurred in the last few years together with their significance, especially in relation to the drive for greater productivity which is going on both here and there. Some twenty visits were made, some of these by part of the team only, in the 42 days' stay. Admittedly the firms visited were chosen, by a consultant in this field, from among those who are using the most modern practices in the control of quality, so that the picture obtained is not necessarily representative of all American industry. There is evidence, however, particularly in the membership figures of the American Society for Quality Control, that these practices are fairly widespread.

Broadened Meaning of the Words 'Quality Control'

To the statistician the most striking development is the wide meaning given to the words 'quality control.' Over here, and indeed there, the name was for long reserved for the statistical technique involved in controlling a manufacturing process by sample inspection at the time of manufacture. The new conception of the meaning and scope of quality control accepts this as one of the tools but applies the name to the wider management function of ensuring that what is made by a manufacturing organisation is what was intended.

This broadened conception of quality control as a new branch of management has been reported before, but it needed a visit to make me realise the full implications of the change. The management of a manufacturing organisation needs, in order to manage, information on costs, on quantity, and on quality. The first two of these have long been generally accepted as essential; the third has usually been left to

* Nominated by the Royal Statistical Society and seconded by the Ministry of Supply.

a section of limited scope working mainly within the production department. What is being done in America and by some firms in the United Kingdom is to lift the question of quality on to a level with the other two, to make it a major concern of management and, in doing so, to study and state the principles that should indicate the methods to be used in each part of the organisation. The result is a more efficient management, a clearer appreciation on the part of members of the management team of what is expected of them, a reduction in the friction which has often been the outcome of the absence of that clear definition, and a product which, in serving the user better, releases capacity and energy on a national scale.

The Scope of Quality Management

Quality control seen in this way as a new branch of management requires to be organised to do at least three things. It needs, first, to put itself in a position to assure top management about the current quality of the products; second, it must inspect in order to prevent faulty articles passing at certain critical stages of manufacture; third, it must pass the information it has obtained to those departments which will use it to prevent further faulty production.

A little thought will show that these functions are akin to those of other branches of management. The costing branch, for example, must be in a position to inform top management of the current level of costs in order that benefit from changes may be assessed. It must continually be making assessment of costs at various points of the organisation and it must inform others of any untoward variation in order that the cause may be removed as quickly as possible.

It is not only within the manufacturing organisation that the scope of quality control has been extended; producers also recognise that the external influences must be allowed full play if the product is to be successful in a competitive field. These influences express themselves sometimes positively as complaints by customers, sometimes less positively by a falling off in sales. Market research may reveal features of design which influence demand; sometimes such features must be sought by field experiment or life test. Whatever their source they cannot operate successfully on the manufacturing organisation unless channels are provided for the interpreted information to flow to the appropriate departments. The quality department is recognised as the appropriate one to provide these channels. Broadly speaking, the operation of these external influences is very similar to that of inspection within the factory, for just as an operator in charge of a machine makes his article, tests or measures it, and adjusts his machine accordingly, so the whole manufacturing organisation must be ready and provided with the means to carry through this same cycle of making—testing—adjusting. The ideal firm will, for example, be constantly measuring the reaction of the market to its product rather than waiting to do so until the sales fall off.

It is this rather broad conception of quality control, a conception which includes outside influences which should be allowed to modify the specification as well as that internal control which ensures conformity with specification, which is characteristic of American practice today. The methods and techniques employed are not very different from those we are accustomed to in the United Kingdom, but the organisation required is different.

The Organisation for Quality Control

This broadening of the field of operation of quality control requires a quality control department to undertake tasks additional to the ones usually assigned to inspection within the factory. But since these tasks are directed to the same end as inspection, namely the control of quality, it is natural for quality control and inspection to be co-ordinated. The results of work in the new field, however, are to be used to modify design and specification, things which should be the firm foundation for the manufacturing part of an organisation. Thus the new quality control department must take its place alongside 'design and methods' and on an equality with them and with 'production.' It is not, however, solely for this reason that the new organisation requires a head in the first rank of the management hierarchy; it is also because top management have also a duty to check the operation of the organisation it has set up, a duty it can delegate to the quality control section but only if that section is directly responsible to it. Thus the need is for an organisation with its head free from the day-to-day difficulties of production yet with parts permeating the whole. Such a requirement is not unique; it is that, for example, of the accounting service and of the personnel department. The manner in which such a functional organisation can exist in parallel with an executive one like the production organisation has been discussed by Mr W. B. D. Brown.[2]

One of the early tendencies when mass-production techniques were new was the separation of the two functions of production and inspection. The two were regarded as incompatible and not to be exercised by the same person. In fact, it has been the deliberate policy of some to set the producer against the inspector so that one could press for greater production while the other opposed this as far as was necessary to maintain quality. Such conflict cannot be productive. There can be no escaping the fact that good quality is only achieved if everyone is working to that end without conflict. The American realisation of this is summed up in the slogans used in the factory: 'Quality is everybody's business' and 'Quality depends on me.' The type of organisation adopted by American management gives back to the production departments the responsibility for quality while still keeping the functions of inspection and production separate. A shop superintendent pressed to keep to a schedule may be tempted to relax his standard of quality in a particular instance; but he is less likely to do so if the exercise of

the inspection function by another person, reporting to him the result of the inspection, leaves him still with the responsibility for quality.

A typical form of the organisation network is shown in the chart of Fig. 1; where there are several plants in the same manufacturing organisation it is even more important that the central administration shall contain a quality control organisation with its own links with the inspectorates of the separate plants.

In these new quality control departments it was interesting to come across the new designation of 'quality control engineer' or 'quality engineer.' Those who employ them are emphatic that they should be engineers who have gained a knowledge of statistics rather than statisticians with a secondary knowledge of engineering. As many as six were encountered in one firm. They were employed on the analysis of the reports on quality from the various sources, on the solution of production problems, and in the instruction of foremen. Not infrequently their statistical knowledge is gained from a short, intensive course at a neighbouring university; it is rare that a man will go to the firm fully equipped in both engineering and statistics.

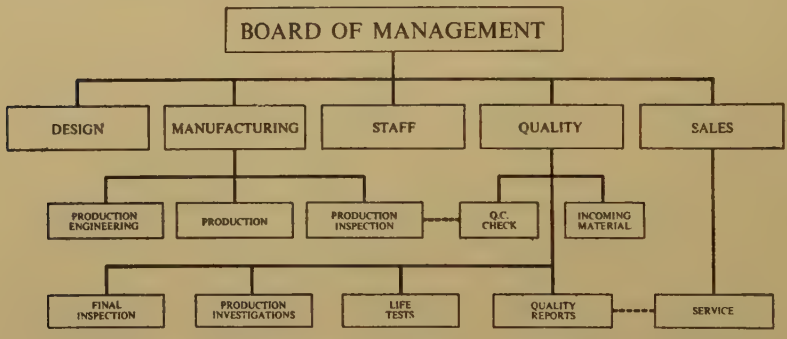


FIG. 1. Organisation chart showing the position of the quality control department.

The Activities of a Quality Control Department

In America, quality control like most other topics has been the subject of many lecture courses and has consequently been methodically dissected and its basic principles stated. This scientific approach is of value, of course, in enabling consistent applications in differing situations. On several occasions the writer was shown how certain parts of an organisation satisfied these basic principles that were being taught.

For example, in each of its fields of activity a quality control department can be thought of as fulfilling in varying degrees the three functions already mentioned, of accepting or passing good articles, preventing bad articles from being made, and reporting on the levels of quality achieved. The reporting forms an assurance to the management that its aims are being fulfilled and to the consumer that the goods are suitable.

Of these three functions it is acceptance that has been predominant hitherto. The others are called into play only by the broader view now taken of the duties of a quality control department. The exercise of the acceptance function makes the department the arbiter of what is good or bad, that is, the interpreter of the specification. Such interpretations can naturally be most effectively made by a department that also has as one of its duties the analysis of faults found at later stages of manufacture and in the working life of the article after it has passed to the user.

The prevention function calls for experiment and investigation into materials and processes which are beyond what may ordinarily be expected of a production department. Long-term improvements in production processes and the tracing of causes of production troubles can also be listed under this function.

The assurance function is given great importance in the newer fields allotted to quality control. The department is thought of as exercising the delegated checking function of top management and so must make reports which form a kind of audit for management. Such reports are in fact essential if management is to keep any control over quality. They can be as detailed or as summary as is desired, but when detailed they have the further use of promoting some rivalry in the field of quality between plants or between shops. For this purpose there is no reason why some general measure, a quality index, should not be used for comparative purposes. If the reports give the amounts of scrap they form a means for management to check efficiency as well as quality. The assurance afforded by the activities of a quality control department extends to the user, and this fact is often used in sales publicity.

The four main fields of activity in which these functions are exercised are (*a*) the inspection of incoming materials, (*b*) inspection during production, (*c*) final or functional inspection, and (*d*) the collection, analysis, and interpretation of service reports, life tests, and customer complaints.

Incoming Materials. Most manufacturing organisations need to keep some check on the quality of their raw material. When the 'raw' material consists of finished parts which are to be assembled the work can assume great proportions, and the help of special methods is extremely desirable both to reduce the cost of inspection and to save the cost of holding and storing between delivery and use.

The inspection work on these 'bought out' goods is often duplicated, since the supplier also needs to inspect in order to control his own quality. Further, the purchaser's inspection must often be carried out on a large scale because information about order of production and state of control of the manufacturing process is not available to him. These two major causes of wasteful expenditure have naturally received attention and the larger firms especially have achieved a large measure of co-operation with their suppliers. This co-operation may consist of a record in the form of a statistical quality control chart covering

production before, during, and after the production of the batch, or of visits by the purchasing company's inspector to see these charts from time to time. The initiative to install the statistical quality control system has usually come from the purchaser, but one supplier met with was in the habit of sending the equivalent of a sample frequency distribution with each batch of goods.

In this field of inspection the acceptance function is obviously very much to the fore, but the new development just described is of the nature of defect prevention. The records kept of each supplier and the remainder of the activity in general form an assurance to the rest of the purchaser's organisation that the articles purchased will be suitable.

Inspection during Production. In keeping with the principle that the producer is responsible for the quality of the work he does, the main work of inspection in this field was found to be carried out in very close touch with the production department. Where there is a strong quality control department actively engaged in the other fields described, it is not unusual for production inspection to be managed by the production department itself with the quality control department acting as advisers and making their own check on the inspection from time to time. In view of the very strong responsibility for quality required to be carried by the head of a production department this is a very good arrangement. The production head can be assured of a rapid redeployment of his inspection force when circumstances require it, and a lot of the old antagonism between 'inspection' and 'production,' especially at the operator level, disappears.

The break which is thus made in the vertical chain of functional authority represented in Fig. 1 need have no adverse effect on standards of quality, since the quality control department still retains a measure of supervision and especially since the final inspection is then usually left entirely in its hands. The production inspectorate gains by having its long-term production difficulties investigated by a staff trained in statistical analysis, a training which is required for the other activities of the quality control department but which might be considered unnecessary for production inspection staff.

Final Testing or Acceptance Inspection. Where the production inspectorate and the quality control department are divorced in the way just described it is usual for the final testing to be carried out by the quality control department, who are then able to report on the quality of the articles actually emerging from the factory. They thus exercise their assurance function to top management besides intercepting any faulty articles before they reach the customer. As receivers of the articles from 'production,' and being independent of 'production,' they are able to take the customer's point of view and report accordingly. The information they can pass back can also be used to improve the quality of production.

According to the nature of the product this stage of inspection is carried out either by sampling or by total inspection. The fact that

this is the last stage before the article reaches the customer leads many firms to make a very thorough and total inspection. If sampling is used it must usually be the rather large-scale sampling used for batches, although in some cases the readily available knowledge about the order of production and about the control during production enables the sampling to be reduced to the proportion needed for statistical quality control charts.

Final testing often consists of, or is combined with, a functional test, especially of such complicated articles as refrigerators, motor cars, and radio sets. The information gained during a functional test is of importance to design and methods departments as well as to the production department, so is more properly the sphere of the more widely based quality control department rather than the production inspectorate.

Outgoing Goods. By the time the manufactured articles are packed and on their way out of the factory there should be very little that has escaped the inspection net. What there is, however, could be important in terms of time and work required to make corrections after the articles have gone out of the factory into the hands of the ultimate consumer. It was not unusual, therefore, to find the quality control department taking a small sample of the outgoing goods, unpacking them, perhaps after simulating the treatment received by the package during a journey, and giving them a thorough examination and test. In addition to checking on what is happening in the packing department much can be learned by such a practice. Where a reputation is at stake it is not uncommon for the manufacturer to make purchases of his own products in the open market in order to make a similar check at a still later stage.

In the Field. The work of the quality control department in the field, that is outside the factory, consists of the collection and analysis of customers' complaints and of reports from the sales organisation and from the service engineers. The recognition of these as potential guides on quality and their channelling through a department having intimate contacts with the departments concerned with design and specification are major factors in reducing the lengths of the periods when the quality is, as it were, 'out of adjustment.'

Statistical Techniques in Use

Bearing in mind that the firms visited were chosen as leading in the latest practices it was not surprising to find sampling methods in extensive use; what was surprising was to hear them sometimes described as new. The history of the development of statistical methods was, of course, well known to the few, but there, as here, must exist many firms to whom they are new or even unknown.

Total inspection was by no means abandoned, but economies had been effected by passing to the sorting room only those batches judged on a preliminary sample to be below a given quality. Incoming stores

in particular were dealt with in this way, the irregular load on the inspecting staff being evened out by supplementing with non-urgent work of a routine character. One firm had gone to an unusual amount of trouble over the randomness of its samples. The inspectors were supplied with lists of random numbers suitable for various batch sizes. They had, too, a table marked out in numbered squares on which a batch of small articles could be spread as a preliminary stratification.

The same firm had developed the 'lot-plot,' a sampling method of delineating a frequency distribution and of calculating confidence limits for the quality of the lot.[3] This 'lot-plot' had been reduced to a standardised routine, so that in addition to its regular use it can be called for by a quality engineer studying a special problem to provide him with the data he requires without having to give detailed instructions.

The Munitions Board has published a Standard[4] containing a selection of sampling schemes suitable for quotation in military inspection specifications. This standard was frequently referred to by firms as the source of their sampling schemes. In fact it appeared that government inspection had played quite a large part in spreading knowledge of sampling techniques; it was recalled, too, that many of the study groups which later formed nuclei for the American Society for Quality Control had their origin in government-sponsored lecture courses during the 1939–45 war.

The extent to which control charts were used was variable. In some plants they were prominently exhibited on almost every machine as an incentive to the worker; in others they were in use only on troublesome processes and might only be used for a limited period. The firms visited were mostly fully aware of the advantage of measurement over go/not-go gauging in detecting changes in any process. The additional cost of the measuring instruments compared with gauges was considered well worth while, though this opinion must of course be considered against the background of their usually large production. On the other hand, go/not-go gauges have to be made specially for each job, whereas sensitive measuring gauges are now standard articles.

The Inspection Specification

Within the factory the word 'quality' can refer only to degree of conformity with specification; the absolute or specified quality of the article being made has already been decided by the manufacturing policy of the firm. Great importance seemed to be attached to the completeness of this specification, since anything lacking in this gives an opportunity for the misunderstandings and deviations which produce the unwanted variations in the final product. In particular, the specification or one copy of it will include the detailed instructions for inspection, the special properties to be looked for, the gauges needed, or the sample size to be used on batches.

Publicity for Quality

It appeared that in most of the firms visited the organisation for controlling quality had been installed to the accompaniment of much publicity within the plant. The opportunity would be taken of a major reorganisation such as occurred at the end of the war contracts to seek out the latest techniques in each branch of management. Having decided on a quality organisation and chosen the head and his assistants the management would let its interest in quality be known in no uncertain terms. One firm was reported to have engaged an advertising firm to plan the introduction of a quality 'programme.' A special number of the house magazine would be devoted to the subject, posters would be displayed, and competitions would be organised for slogans. More important still, some such evidence of interest on the part of management was planned to be continued.

At the back of this policy of publicity was the realisation that the inspection and quality departments by themselves cannot produce the desired result. Quality must ultimately be the responsibility of the production department, of the individual operators; all that inspection can do is to assist by providing the information. Its function of refusing to pass on the bad-quality goods does not in itself produce quality. The quality publicity was therefore, with reason, directed at the operators and all directly concerned with production.

It was not as a rule found feasible to relate bonus schemes to quality, though group bonus schemes might be so related. The difficulty was that so many operators were involved in the production process that any assessment of blame involved much argument and it was considered worth while avoiding the consequent friction by basing the bonus only on group quality.

The American Society for Quality Control

The real evidence of the upsurge of interest in quality in manufacture which is taking place at this time in America comes from the activities of the American Society for Quality Control.

Nothing quite like this society exists in the United Kingdom. Membership is open to all interested in the activities of the society. It issues a monthly magazine, *Industrial Quality Control*, available only to members and to libraries, containing news of section activities and members together with articles on the subject of quality control. The membership is over 4 500 in 53 local sections, including one in Canada and one in Mexico. These local sections were formed as a result of the government-sponsored lecture courses on statistical quality control given in the industrial centres during the war. In many cases a discussion group continued as a local quality control society; it was not until 1946 that a move was made to unite them into one society.

More recently a decision had to be made whether or not to become a section of the American Statistical Association. It is significant that the decision not to do so was made on the grounds that the subject of

quality control is a wider one than the statistical techniques known by that name and that to join forces with an avowedly statistical body would restrict the membership. The decision has been justified by the large increase of membership which has resulted from a statement of these wider aims, and by the large attendances at section meetings and society conferences. It is evident that, all ranks of industry being eligible to join, a new source of membership has been tapped.

Many of the sections or local chapters of the society offer instruction, in the form of either 'clinics' held just before chapter meetings or an instruction course preceding a conference jointly arranged by several sections. At a 'clinic' members' problems, usually matters of actual experience arising from their work, are discussed by a panel of the more knowledgeable members of the section, often including a member of the teaching staff of the local university.

I attended one session of a 'two-day intensive training course in elementary statistical quality control' arranged to precede the Sixth Mid-west Quality Control Conference of the fourteen mid-west sections of the ASQC. For this a large ballroom of a Chicago hotel had been equipped with a public address system and desks for the 120 or so students. A printed manual gave full notes and diagrams for the twelve lectures given by lecturers drawn from the four universities of the area. The active co-operation of the Chicago Association of Commerce and Industry in both the conference and training course had no doubt done much to bring about the large attendance. The time-table and titles of the sessions are given to show the scope of the course.

Monday

9.00– 9.50	What is statistical quality control ?
10.00–10.50	Interpreting industrial data
11.00–11.50	Construction of average and range charts
1.30– 2.20	Using an average and range chart for process control
2.30– 3.20	Frequency distributions and control charts. Process *vs* product. The normal distribution
3.30– 4.20	The control chart for fraction defective—what it is and where it applies

Tuesday

9.00– 9.50	Using fraction defective charts in industry
10.00–10.50	Other attribute charts and their use in industry (number defective, defects/unit, and average defects/unit)
11.00–11.50	Fundamentals of acceptance sampling
1.30– 2.20	How and why modern sampling plans work
2.30– 3.20	Choice of sampling plan—single, double, multiple; how they compare
3.30– 4.20	Available standard sampling plans: Dodge-Romig, Military Standard 105A

It was not clear what class of person would benefit from such a course. It was much too condensed to serve as the training for the quality engineers referred to earlier; yet it might serve as a means of recruiting to this new profession, since many, though already made curious by hearsay, might not without some short introduction such as this ever be tempted to go more deeply into the subject.

This new profession of quality engineer cannot yet be said to be firmly established. The larger firms who can afford to have men engaged full time on problems of quality employ them under this name or as 'analysts,' but as far as I could ascertain they are not trained as such but as engineers with a short added training in statistics, often received after joining the firm. The fact that the name is used is, however, a strong indication of the need for men with this compound training.

Conclusions

As a result of making such a visit as this and seeing for oneself the enthusiasm with which the task of controlling quality has been taken up by some of the larger firms one is inclined to say that the subject is comparatively neglected over here. Second thoughts, however, incline one to the view that the neglect is apparent rather than real. Here and there in the United Kingdom are to be found firms applying these principles, in part at least, and deriving much benefit. What is missing is the methodical study and enthusiastic application by the younger staff which come from discussion, lectures, and the exchange of experience. This may be the reason why more firms are not found following the lead given.

In its current form, quality control is a management subject even though its techniques may be statistical. Enough has been said above to show that successful application does not necessarily require students of statistical method; in fact the participation of practitioners in all branches of management is required. It follows that the opportunity for and the encouragement of this exchange of experience must come not from a statistical body but from one interested in management and whose membership is not limited to any one class of worker. In fact the question arises 'Should there be a British Quality Control Society?' The answer must come from industry itself.

REFERENCES

[1] *Inspection in Industry*. British Productivity Council, London, 1953.
[2] BROWN, W. B. D. *J. Inst. Labour Management*, January 1946.
[3] 'The Lot-Plot Method of Acceptance Sampling by Variables,' *Industrial Quality Control*, July 1950.
[4] *Sampling Procedures and Tables for Inspection by Attributes*, Military Standard 105A. Government Printing Office, Washington (agent, HMSO, London).

AN EXPERIMENT IN THE SAMPLING OF AGRICULTURAL RETURNS

HENRY PALCA

Ministry of Agriculture and Fisheries

This article describes a sampling experiment on the agricultural returns of June 1951 in which national totals for labour, acreage, and livestock items were estimated from one-half of the parishes. Various methods were used for making the estimates, and the results are compared with actual totals to give guidance for future procedure.

Introduction

The agricultural return which farmers in England and Wales fill in every June (on acreage, livestock, and labour numbers) is the most important of the year, since by that time the cropping programme for the coming harvest year is fixed for all the main items. There are some 385 thousand holdings altogether, from 12 thousand separate parishes, and from the end of June onwards the Ministry's Collection of Statistics Branch at St. Annes is sent returns from the local centres in bundles of parishes; the last parish is not usually received until the end of July. In 1951, as an experiment to save time and man-power, the preliminary results were estimated from one-half of the parishes and were available for publication a fortnight earlier than usual. The figures for the remaining parishes were used in producing the final results. As mentioned below, it was the parish and not the individual holding that was actually sampled. Thus the sample has strictly to be considered as one of six thousand separate sampling units.

As a guide to procedure for subsequent years, different ways were tried of selecting the sample, and also of using it to estimate the complete totals, here termed 'raising' the sample. The various estimates were checked later in the year against the results from all parishes. The comparison was actually made with what may be termed the '100 per cent. provisional results,' which correspond to the preliminary results published in earlier years, and which always show certain differences from the final results, partly because of the late returns for some holdings, for which estimates are inserted in the earlier totals, and partly because of later corrections to individual returns. It was therefore fairer to use these figures rather than the final ones for measuring errors due to sampling alone.

It is not the purpose of this article to examine the saving in time and labour due to the new procedure, but rather to discuss the reliability of the results and to see how far theoretical improvements to the

sampling and raising methods were in fact of more than academic interest.

The Size of the Sample

At first sight it might appear that the sample was unnecessarily large, and that, say, a one-in-ten sample might have been sufficiently reliable. It might even be wondered whether a one-in-two sample was worth the trouble of sampling at all. These objections are certainly valid for many types of survey, but they do not apply to the June agricultural census here considered. The results from this comprehensive census are important not only for national totals, but also right down to the parish totals (records for which date right back for 85 years and are frequently used by research workers in agricultural economics and statistics) and even further down to the details for each individual holding, since it is essential for present administrative and advisory purposes to have an up-to-date record for every farm in the country. For these reasons any early sample of the June returns can only be the first stage towards a complete enumeration, and the uses of sampling are confined to the above-mentioned purposes of getting speedier preliminary results and of saving money by spreading the work of full summarisation over a longer period in the year. The interval of time between the preliminary and the final results is as long as four or five months, partly because of the lateness or incorrectness of some returns and partly because other censuses intervene. The preliminary results have therefore to be used for quite a considerable time.

The June return contains some items which have very small totals even for the country as a whole, but which are nevertheless important; and others which, although relatively large in total per county, are still distributed very irregularly from farm to farm—for example in Norfolk on one occasion 10 per cent. of all the ducks were on one holding.

Figures for separate counties are required also at an early date by County Agricultural Committees. The pre-sampling stratification (i.e. sorting) by sub-districts within counties—which was adopted in this experiment—was not enough to keep all county errors low; for many items only a very complicated system of stratification, combined with the use of 'optimum' sampling fractions, would reduce errors of such totals to below the order of magnitude of actual year-to-year fluctuations, even with a 50 per cent. sample. For example, as many as two-fifths of the Rutland totals were found to have errors of over 10 per cent. when raised by the best method, the totals in such cases being mostly very small. In practice, with this necessarily simple sampling scheme, it was only considered safe to publish provisional results for the more important items, or combinations of items, in even the large counties, and this caution was found to have been fully justified.

On the other hand, England and Wales totals were well within the limits of required accuracy, and had these national totals alone been needed a much smaller sample would have been adequate.

The Sampling Procedures

Since the completed returns were received in batches of whole parishes it was convenient to treat the parish and not the individual holding as the sampling unit. The sample comprised about one-half of the parishes in each district of a county. Every holding in a selected parish was used. Four different sampling procedures were applied to separate regions of the country, two National Agricultural Advisory Service provinces to each.

I. *Random Sampling* (East and West Midland Provinces). Parishes were selected at random for each National Agricultural Advisory Service district (a sub-division of the county), first dividing up the district into pairs of adjacent and similarly-sized parishes to ensure an even distribution over the district, and then choosing one parish at random from each pair, using a table of random numbers for this purpose.

II. *Acceptance Sampling* (South-West and Northern Provinces). Those parishes were taken which happened to reach St. Annes first, covering about one-half of the crops and grass area of each district.

III. *Mixed Sampling* (*A*): *Random-Substitution* (South-East and Yorkshire and Lancashire). Parishes were chosen at random from pairs exactly as for method I, but for this method a strict time limit was established for each district. As substitutes for those sampled parishes not received by the specified date, it was intended that the non-sampled member of the relevant pair would be used whenever it had arrived in time; otherwise another non-sampled parish was to be used, as close and as near in size as possible to the missing parish. The details for this method are recorded here for completeness, but in fact all the sampled parishes arrived before the time limit had expired, so that the method used for these two provinces was method I, and is treated as such in the ensuing discussion.

IV. *Mixed Sampling* (*B*): *Random-Acceptance* (East Midland and Wales). Here two-thirds of all parishes in a district were selected at random, and from them the first half of the district received was taken as the sample.

It seemed reasonable to expect that results would be most reliable for method I, and next for IV and II respectively, while costs and speed might be in reverse order.

Methods of Raising the Sample

(*a*) *Raising Item by Item.* Within each county the totals for the parishes entering into the sample for each of the main items on the June 1951 Agricultural return were compared with those for the same parishes in June 1950. The percentage changes as between the two years were then applied to the full county results for June 1950 in order to estimate the respective county totals for June 1951. This was

the method used in presenting the preliminary results, as it was likely to provide more accurate results than the simple raising procedures (*b*), (*c*), and (*d*), which were tried out on the same material afterwards to test them against (*a*).

It should be added that the June 1951 total crops and grass area for each county was obtained independently for all parishes, as this was a relatively simple procedure from another source, and the sub-items for crops and grass were adjusted proportionately to agree with the respective totals. Rough grazing areas were also obtained from this source, and do not therefore form a part of the sampling experiment, except as a tested raising factor for sheep.

(*b*) *Raising by Total Crops and Grass.* Within each county the ratio of total crops and grass area in the county to the total for parishes in the sample was applied to every item on the returns.

(*c*) *Raising by Sub-group Totals.* The crops were all raised by the total tillage ratio, grasses by total grass, and each labour and livestock item by the respective sub-group total. Thus for cows and heifers in milk the sample was raised, county by county, by the ratio of June 1950 total cattle to the June 1950 total cattle for the sampled parishes only. Once again adjustments to acreages were made from the full crops and grass totals.

(*d*) *Raising by Doubling.* The sample totals for each item were simply doubled, and no adjustments at all were made.

After (*a*), results were expected to be most reliable for (*c*), (*b*), and (*d*), in that order.

It will be noticed that, had time allowed, it would have been a slight improvement for the first three methods to have been raised by each county's districts separately.

Comparison of Sampling Methods

The three sampling methods—random, random-acceptance, and acceptance—are not easy to compare, since it was impracticable for any region to be sampled by more than one method. No very obvious differences leaped to the eye, but a series of paired comparisons between the two extremes, random and acceptance sampling, suggested that when results were raised by method (*b*) (by crops and grass throughout) the random sampling procedure was worth the extra trouble. When results were raised item by item, however, there was no difference to be observed between the reliability of the sampling methods.

The simple explanation for these results would be that any bias in the acceptance sample is sufficiently corrected for by raising from an identical sample. Indeed, to judge from an analysis of percentage positive results, the acceptance samples raised by (*b*) were a little biased towards low estimates, and this seems to have been because parishes with larger holdings in them were a trifle quicker in reaching St. Annes. For many items, density per acre varies indirectly with size of holding,

whereas there is no evidence that annual proportional changes are similarly correlated with size.

Comparison of Raising Methods

The errors in county results from raising by methods (*a*) and (*b*) were compared for each item, and in every single case the majority, usually a large majority, of the counties were found to have more accurate results by method (*a*) than by method (*b*). The acreage items showed a smaller gain than the other items, measured by the size of the percentage of total comparisons where (*a*) is better than (*b*), but this is probably because the former are anchored under both methods to the known totals of crops and grass area, so that there is less room for manœuvre. The over-all result was that a good two-thirds of the cases showed (*a*) to give lower errors than (*b*).

On comparing the three sampling methods, the acceptance method seemed to gain most from (*a*), which links up with the findings of the previous section; but the random-acceptance method did not consistently adopt the intermediate positions, and this suggests that the choice of raising method was a more powerful influence on accuracy than the choice of sampling method.

The sheep items were the weakest for method (*b*), but they are improved markedly when the raising factor includes the area of rough grazings. Even then (*a*) is superior.

At first sight it appeared surprising to find that it was not always those items likely to remain on the same farm each year which enjoyed the greatest advantage from method (*a*) raising. One reason for this must be that the county totals for some of these items (e.g. horses) are in fact very small, and an analysis confined to small items showed that the advantage of (*a*) over (*b*) became negligible in such cases. This is presumably why the sub-item totals (e.g. total cattle) showed in every case a stronger advantage in (*a*) over (*b*) than do the sub-items themselves. Another reason for the disappointing results for these relatively fixed items is that there is an overlapping influence at work, namely the effect of annual changes upon the relative efficiencies of methods (*a*) and (*b*). Some of these 'fixed' items are increasing or diminishing at a rate above the average, and in an analysis confined to national totals of over one hundred thousand (to eliminate the effect of smallness of item) it did appear that where the movement over the year had been sharpest then (*a*) tended to lose its advantage over (*b*).

Other raising methods have not yet been mentioned, but results for them were of less practical importance. Method (*d*) was unimportant because its expected inefficiency was so clearly shown for every item. This method is the extreme case and for that reason no adjustments at all were made, for example raising by holdings ratio or even tying down crops and grass items to the known totals. Method (*c*) gave results only a little less reliable than those for (*a*). It is doubtful whether the extra work of raising by nearly 80 separate ratios per

county, as compared with only 8 under (*c*), would be really justifiable for the grass, labour and livestock items, but it does seem to be so for tillage, presumably because this definition covers a wide variety of items such as wheat, flax, fruit, and bare fallow.

The Size of the Errors

The above comparisons have only shown that in this experiment some of the theoretical advantages of certain sampling and raising techniques were strong enough to reveal themselves in action. All mention of the actual errors obtained has been left to this concluding section.

The great majority of the totals for England and Wales raised by method (*a*) have sampling errors of less than 1 per cent.; among them are those of all the most important ones, e.g. main cereals, potatoes, dairy herd, sows, ewes for breeding. The only ones to show sampling errors of any size are young men employed on a casual basis (6·5 per cent. on a total of ten thousand), ram lambs (5·8 per cent. on 68 thousand), and heavy stallions (12·8 per cent. on only 500).

For most of the items with errors of over 2 per cent. the difference due to sampling error is very small compared with the difference between the actual June 1950 and June 1951 totals. Ram lambs are an exception to this: unfortunately the positive errors tended to be clustered in counties with the larger number of ram lambs (especially Wales), but there was no evidence of systematic bias, since in all there were 29 negative to 30 positive county errors, and these were symmetrically grouped. There seems therefore to be no reason to fear that the results for any other year would be so astray, although they may be of the order of 5 per cent. because frequency analyses show this item to have a comparatively high sampling error.

The error of nearly 13 per cent. for heavy stallions represents an absolute amount of less than one hundred. Most of the errors could be accounted for by two counties, and if their results had been exact then the national error would have been under 5 per cent.—an example of the precarious nature of estimates for such small items.

There are two items, namely fruit and vegetables (not for sale), and other crops, which are subject to quite considerable adjustments in the final results, a reminder that even the old '100 per cent. provisional results' were never immune from error. In fact the results from the 1951 sample are for some items nearer to the final 1951 results than the 1950 100 per cent. provisional were to the final for 1950.

In the published results for individual counties raised by method (*a*) certain items were amalgamated, and any items of less than one thousand (acres, labour, or livestock numbers) were omitted. Even so, one-eighth of the results had errors of over 5 per cent. and one in twenty-five were over 10 per cent. in error. In a normal distribution with standard deviation of 5 per cent. one would find one-third of the results to be beyond the 5 per cent. range and one-twentieth beyond 10 per

cent. Regarding the errors as a whole, one might therefore be tempted to say that the sampling errors of published county figures obtained in this way are rather less than 5 per cent. This over-all distribution of errors has, however, a very abnormal shape, mainly because there is a highly irregular range of item sizes per county, so that the above general statement would really be meaningless.

The size of the item per county does indeed influence errors. Items of between 1 000 and 10 000 units were at least 5 per cent. in error once in five times whereas items of over 10 000 units were as far wrong only once in fourteen times.

Of all the different categories of items, those for sheep are the most prone to error, and county totals for any of them under 10 000 should not be regarded as safe. The sheep population in many counties is clustered on a relatively small proportion of all farms in the area.

Indeed, the number of farms in the county which have a particular item on them is most probably the chief factor in determining sampling errors and it may bear little relation to the total number of farms for some rare or peculiarly distributed items. This clustering (or heterogeneity of farming types) may be more marked in some parts of the country than others; the Eastern province seemed to suffer from it rather badly. The weak county items are by no means automatically so for England and Wales as a whole—for instance the England and Wales error for mixed corn (a bad county item) is negligible. The good county items can, however, be relied upon to be good for the respective national totals too. Generally speaking, the errors for an item on a county, province, and national basis were found to be in about the ratio $1 : \frac{1}{3} : \frac{1}{8}$, or what might be expected from their respective average sizes.

How far could these errors have been reduced by more refined sampling techniques, had it been practicable to make use of them? A reply to this is provided from analyses for certain counties of frequency distributions; among them may be mentioned those for Hertfordshire (1939) given as examples in Dr Yates's book *Sampling Methods for Censuses and Surveys* (Griffin, 1949).

To begin with, it would have been better to use the individual holding and not the parish as sampling unit, but this improvement is not by itself very marked, for, roughly speaking, errors of an unstratified random sample of holdings, using no supplementary information for raising, seem here to be of the same order of magnitude as those from a random sample of parishes raised by crops and grass. In fact a random sample of parishes selected with probabilities proportional to size of parish would perhaps have shown a more pronounced improvement, though the complications involved in selecting and raising such a sample would have been impracticably great.

The use of the holding as the sampling unit would, however, enable the population to gain the most benefit from further stratification. The most obvious scheme of this type would be to stratify by size of holding.

For some items where the average size of item per holding varies strongly with holding size (e.g. wheat, regular labour) there is a marked reduction in sampling error, the method having an efficiency, measured in terms of the respective variances, of one and a half to three times that of a random unstratified sample. For other items, such as some of the rarer crops or the poultry items, the gain in efficiency is negligible.

A further substantial reduction in error is obtainable from the use of a sampling fraction varying according to the standard deviation for each size group. This gain is even marked in many cases where the use of a uniform sampling fraction showed little advantage over an unstratified random sample. Unfortunately the optimum sampling fractions (i.e. those which would minimise the sampling error for the item in question) differ enormously from item to item, and a general sampling fraction which minimised over-all errors, or which gave low errors for certain types of item only, would result for some items in higher sampling errors than those obtained from a uniform sampling fraction scheme, or even in some cases than those from a random unstratified one.

There are ways of escaping from this dilemma, either by further stratification of specialist types of farming, most effective for certain crops grown on relatively few holdings, or by linking results to those for a previous census. This latter method, from an unstratified sample, has been shown in some instances to give sampling errors roughly intermediate between errors from a uniform sampling fraction procedure and those from an optimum set of variable sampling fractions from populations stratified by size of holding. Nevertheless, for the rarer or more changeable classes of item it must always be considered a second best to careful stratification and choice of sampling fractions whenever these latter methods are practicable.

Conclusion and Acknowledgments

This investigation, by revealing the sizes of actual sampling errors, should serve as a forecast of those which might be expected from similar procedures in the future; and at the same time, by demonstrating the satisfactoriness of the results on the national level, may help to create confidence in sampling methods applied to agricultural returns for samples of a smaller and more manageable size.

The author is indebted to the Ministry of Agriculture and Fisheries for permission to publish results obtained in collaboration with his colleagues, in particular with Mr F. W. German, and Mr W. H. Newsome and his staff at St. Annes.

A TECHNIQUE FOR STANDARDISING MASSED BATTERIES OF CONTROL CHARTS

STAFFORD BEER

Mr Beer describes a simple apparatus and a method of standardising control charts, with the aid of which a large number of charts can readily be kept under surveillance. He describes an application to production control in a factory.

Introduction

When control chart procedures have been adopted to provide statistical surveillance of some wide scheme of control, the user is confronted with a number of operational difficulties. Suppose that his control scheme covers fifty measured quantities, each with its underlying frequency distribution and each distribution having parameters peculiar to itself. According to orthodox methods, there will be two charts for each distribution: one of sample averages and one of sample ranges. Since the positions of limit lines depend on the standard deviations, there will be fifty pairs of graphs, mostly with different limits.

Attempts to plot these graphs adjacently on a large sheet of paper are foredoomed to failure. Not only does the variation in spread give the chart a chaotic appearance, but it continually needs alteration. A change in the facts of the situation being controlled can easily mean a change in the mean and variability of a distribution. The limit lines on two graphs must then be redrawn. At best, this will make the sheet of graphs look increasingly muddled; at worst, the graphs may begin to overlap.

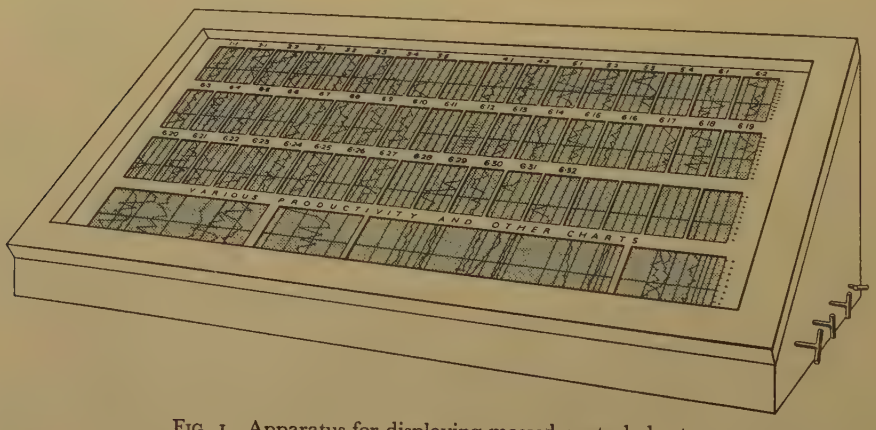


FIG. 1. Apparatus for displaying massed control charts.

Some technique is thus required which will overcome these difficulties and produce an intelligible massed battery of charts. A simple solution is here put forward, with a description of the situation in which it has already been applied.

Apparatus

Fig. 1 shows the apparatus which was devised to house a massed battery of standardised control charts. This is simple both to use and to consult. Here fifty-one similar control charts are identified by numbers (e.g. 6.30); the lowermost line of charts carries additional information about the control system.

Each line of charts is drawn on a continuous roll of millimeter graph paper (70 cm wide) and is attached to rollers operated by T-shaped handles protruding on both sides of the device at its base. Thus the control charts are continuous for a number of years, and move below the fixed mask on which the designations of the graphs are written. This introduces an unfamiliar feature: the time scale runs vertically instead of horizontally.

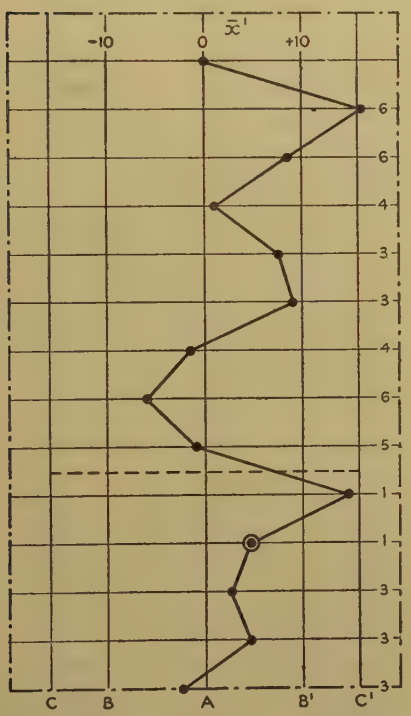


Fig. 2. Statistical control chart for group 6.27. The limits of this standardised chart do not change with changes in the parameters of the controlled distribution.

The single graph which is reproduced in Fig. 2 demonstrates how this works (the numbers on the right should be ignored for the moment). The mean of the group (the quantity measured is productivity) is represented by the central line A; the lines B and B′ represent 1 in 40 limits, C and C′ represent 1 in 1 000 limits. Each point on this graph is the mean productivity for a sample period, and a horizontal line denotes the year end. The 'window' in the mask through which any chart is seen is cut to make visible a year's record of the chart, which is rolled up every time a fresh point is added.

The chief convenience of the charts lies in the fact that they are all drawn on the same scale, with the limit lines at the same distance from the 'mean' line for each chart. The next section shows how this is achieved.

Control of Sample Averages

In the usual control chart for sample averages, the sample average, $\bar{x}$, is usually drawn on a chart with the control level drawn to represent

m, the estimated population mean, and control limits drawn at distances $t\sigma/\sqrt{n}$ above and below m, where σ is the estimated standard deviation of the individual items in the population, n is the number in the sample, and t depends on the probability corresponding to the limits, being 1·96 for the 1 in 40 limits and 3·09 for the 1 in 1 000 limits. For different charts the levels and limits will be at different positions according to the values of m, σ, and n.

In our scheme the sample averages are standardised by calculating for each value:

$$\bar{x}' = \frac{10}{1{\cdot}96} \times \frac{\bar{x} - m}{\sigma/\sqrt{n}}$$

$$= 11{\cdot}4\,\frac{\bar{x} - m}{\sigma} \text{ for samples of 5}$$

Then the control charts for $\bar{x}'$ have the control level at zero, the 1 in 40 limits are 10 graphical units away from the control level, and the 1 in 1 000 limits are 5·75 graphical units further away, whatever the values of m and σ. A change in either m or σ changes the calculation of $\bar{x}'$; it does not alter the chart. When the sample averages are a computed ratio such as productivity, the transformation to the standardised form may readily be incorporated in the computation schedule.

This method involves sacrificing the representation on the chart of any change in the population mean or standard deviation. This can be remedied to a large extent by annotation, and the individual statistician has to decide whether the whole procedure presents enough information.

Control of Sample Ranges

The next question to be discussed is that of sample ranges, and here it is hoped to justify a more radical reduction in the customarily presented information. Orthodox range charts can conceivably be treated in the same way as the charts for sample averages by plotting the range on a standardised chart such as that already discussed, by using a statistical transformation. This would, however, be a more complicated method than the situation usually warrants, and a simple substitute of the kind described below is often sufficient.

I take the general point of view that the curtailing of orthodox statistical procedures is permissible when strictly limited aims have been defined. Cases are sometimes noted where statistical effort is disproportionate to the accuracy of available data on the one hand, and to the nicety with which subsequent action can be graded on the other. Orthodox charts of sample ranges seem frequently to fall into this second category. In the present context, certainly, an exact estimate of the probabilities attaching to sample ranges is totally unnecessary.

If the line for averages should jump outside its control limits the range chart is usually consulted to judge whether the displacement is a

uniform change and therefore serious or the result of wildly erratic behaviour in perhaps one individual in the sample. Conversely, if the average chart is quite steady the range chart is consulted to ensure that this appearance is not illusory and merely due to the cancelling of extreme values of individuals within the sample. Thus it is submitted that, at least for the routine use of a massed battery, the statistician is only interested in *zones* of the range chart, such as the following.

In an ordinary control chart for ranges we may use five limit lines; one representing the mean range, which runs between two pairs of asymmetrical limits. These five lines embrace four 'zones of probability.' A fifth zone will lie between the lower outer limit and the base line of zero range, and a sixth zone will exist above the upper outer limit. If it be conceded that a range value can, in exceptional circumstances, be so high as to be absurd, a seventh zone will also arise. For use in massed batteries these seven zones may be coded as in Fig. 3 and the zone numbers may be written against the points plotted to represent sample averages. This will indicate each point's significance in relation to the sample range. The 'absurdity' value has been found useful as indicative of invalidity in the sample, perhaps due to a clerical error, rather than of lost process control.

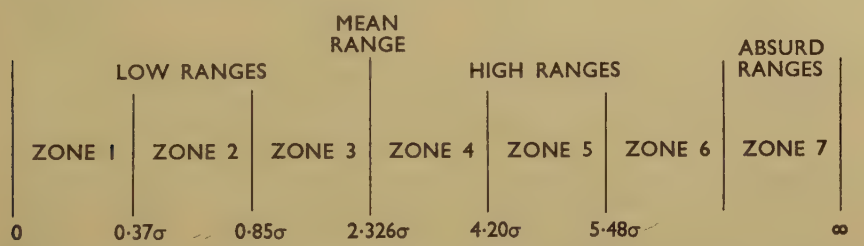


FIG. 3. Zones and boundaries for range chart in samples of 5. The boundary between zones 6 and 7 is chosen arbitrarily at some high value.

A Practical Application of the Massed Battery of Charts

A scheme of production control had been installed in a works department producing a great variety of products on varied machinery. The essence of this scheme lay in its ability to forecast production times in circumstances predictable only in terms of statistical chances.

It was certainly possible to compute, in any given case, what the theoretical production time should be, and a multi-stage nomograph was designed to do this. But the contingency factors of variable efficiency in mechanical performance and materials handling, of arbitrary hold-ups and breakdowns, and of inhibited levels of activity, could not be accounted for in this way. A productivity index was therefore calculated to measure the effect of these losses, which consisted of the percentage ratio between theoretical and actual times for a given job.

This index was computed for every production job done over a very long period and the distribution of the indices was examined. With

the aid of machine methods to sort punched cards bearing these data, various statistical analyses were carried out and causes of variation in productivity were identified. Thus it became possible to divide production into forty-five significantly different groups. Each group was verbally defined in terms of machines and processes and was associated with a definite frequently distribution of productivity indices, estimates of the parameters of which had been calculated.

A final stage could now be added to the nomograph. This multiplied the theoretical time for a job by the reciprocal of the mean productivity appropriate to the group to which the job belonged. Thus a dependable forecast of actual time became obtainable from the nomograph in a few seconds' work. Forecasts predicted in this way have been used for all purposes of production control for a year, and have proved very accurate.

It is, however, evident that this system can only operate in this rigid form while productivity remains substantially constant. If the factors affecting productivity change the nomograph ceases to be appropriate. Hence it is essential that the mean level of productivity for each group should be constantly watched, and this is done by routine sampling from each group coupled with the use of control charts. When it becomes clear from the chart that a significant change in mean productivity is occurring a special analysis is made of the group affected. Should a new mean be established the movable final scale on the nomograph is altered. In this way all forecasts made take into account the very latest information about changes in productivity on the shop floor.

The working of an actual chart can be illustrated from Fig. 2. The figures on the right indicate the zone of range probability associated with the sample average plotted. (This is an actual example: it figures in the apparatus as group 6.27.) In more than half the cases the zone is either 3 or 4, which is to be expected if control is good. The three cases of zone 6, however, certainly indicate the need for investigation into process variability.

In two instances alone did the sample average point give rise to doubts. On the first occasion the range zone was 6. This indicated that the process variability had gone out of control; hence the limits for the sample average had become temporarily inappropriate and its significance was therefore doubtful. A check sample gave a similar result. Enquiry in the works revealed that the process was undergoing experimental modification in an attempt to improve performance without any definite result. No action was taken. The next sample showed the same process variability, zone 6, with a return to normality in the average. The attempt to improve productivity in this way was, in fact, proving abortive and was abandoned. Subsequently the range came back into control, as can be seen from the chart.

When next the inner limit for the average was exceeded, however, the range was in zone 1 and thus significantly low. This indicated that the upward change in mean productivity, highly significant in its

extent as the chart shows, had been achieved with remarkable uniformity. A full-scale experiment was therefore instituted immediately and a new and significantly different estimate of mean productivity was obtained. This new value was then fed into the production control system; the nomograph was adjusted to give improved forecasts, and the new mean was substituted in the control chart procedure. Thus the next sample average point was ringed to show that the group was now working on a new level of productivity. The ringed point, and those points following, had in fact been calculated from the new figure.

Conclusion

If, as is to be hoped, the use of statistical control is becoming more extensive, it seems likely that large-scale schemes of control may become less rare. In the present case the scheme quoted was part of a pilot system of production control installed in one department. Its success is to lead to a much more general application. The implied problem of statistical control is vast, as the use of several hundreds of control charts must be envisaged. Anyone responsible for such a scheme who tries to keep a personal check on graphical progress will be confronted by a serious problem in terms of time and convenience. The present technique of massed batteries is seen as a first step towards the solution of this problem.

Thanks are expressed to David A. Hopkins, who collaborated, and to the management of Messrs Samuel Fox & Company Ltd for permission to publish this account.

AN INDEX OF THE VOLUME OF NET EXPORTS

WLADYSLAW Z. BILLEWICZ

Polish University College, London

Mr Billewicz argues that for some purposes it is better to measure changes in net exports, that is in exports less the imports required to produce the goods exported, than to measure changes in gross exports. He derives an index of net exports and gives results for the years 1948–52.

THE index number of the volume of exports published by the Board of Trade is 'designed to show movements in . . . exports after eliminating variations due merely to price changes.'[1] It shows relative movements in the volume of exports and reacts to changes in quantity irrespective of what is being exported. This is, of course, as it should be when the interest is in the total volume of exports, but there are problems for which the movement in the volume of the 'British element exported,' that is of *net exports*,* is more relevant. For example, if one wants to compare changes in the real national product with those in exports, or to study exports in relation to the balance of payments, it is the volume of net exports that is important. Further, an index measuring changes in the volume of net exports may be useful in the preparation of certain economic plans and documents, such as the Economic Survey.

It can easily be shown that the gross (Board of Trade) and net indices may move at different rates and even in different directions. The problem is, therefore, whether in practice the divergence between them is substantial enough to be worth taking into account, and whether for the post-war years the gross volume index is above or below that measuring changes in the volume of net exports. If the gross index is below the net index this is evidence of a shift in the composition of exports towards low import-content goods, and *vice versa*. The implications of divergence between the two indices may be seen from the following considerations. From the point of view of international payments the main interest is in net exports, since imports for home consumption (including capital formation) must be paid for by the net exports and the surplus of other earnings over the remaining expenditures abroad. If, therefore, the volume of net exports grows more quickly than that of gross exports, indicating a shift of emphasis towards exports with low import-content, then—whatever the price changes—(*a*) more is gained on exports than could have been gained

* *Net* exports are equal to the value of *gross* exports (as declared) less the value of all imported materials and services consumed in connection with the production for export, that is less the import-content of exports.

otherwise, and (*b*) the volume of imports required to produce exported commodities is reduced. The reverse argument applies when it is the gross index which is the greater of the two.

Thus the net index can throw some light on the arguments used in the current controversy between the 'export or die' and the 'import less and live' trends of thought. Its movements can indicate whether the expansion of exports is going in the right direction. Clearly, exports of cotton goods are not as beneficial as, for example, those of glassware and pottery—a distinction which the net index is, while the Board of Trade index is not, capable of showing. Considerations such as these may lead to a policy of deliberately increasing some types of exports and curtailing others and may thus help to restrict what might be called 'unprofitable imports.'

Choice of the Formula

As mentioned above, the two indices need not change at the same rate. Indeed, they cannot be expected to do so for reasons which become clear when one considers the factors influencing their respective movements.

The gross index is defined as $I_G = \Sigma(p_0q_1)/\Sigma(p_0q_0)$ and the net index as $I_N' = \Sigma(p_0q_1 - P_0Q_1)/\Sigma(p_0q_0 - P_0Q_0)$, where p_0 and p_1 are prices of exports in the base and in the current year, q_0 and q_1 are quantities of exports in the two years, and P_0 and P_1 are prices and Q_0 and Q_1 are quantities of imported materials used in the production for export in the two years.

If there were no import content, then the gross and net index numbers would be identical. As it is, the net index depends on two factors which do not enter into the other index, namely on changes in the import content of individual commodities and on changes in the composition of exports as between commodities of high and low import-content. Thus the net volume index takes into account differences in the import-content of different commodities as well as in the same commodity between two different dates. As a result of the former, the net index tends to reduce the effect of changes in exports of commodities burdened with an import-content greater than the overall average, and *vice versa.*

The expression for I_N' is known as the Geary formula,[2, 3] and it may be useful to discuss its characteristics with reference to our problem. Theoretically it takes into account variations in qualities and types of products and materials because each variety can be treated separately. It also allows for changes in the amount of processing applied to imported materials, since their cost will, in general, vary in proportion with this. Finally, it can be computed as it stands for a single export group or for all groups taken together.

However, the practical difficulties of its application are considerable, since it is difficult to imagine how a series of, say, yearly estimates of net exports could be obtained on the basis of available information.

Another drawback, perhaps not very important in practice, is that if there are violent changes in the composition of both products and materials the numerator of the formula may become negative.

To overcome these difficulties the formula may in practice be approximated by

$$\frac{\Sigma(p_0q_0 - P_0Q_0)q_1/q_0}{\Sigma(p_0q_0 - P_0Q_0)}$$

This approximation involves the assumption that the quantities of materials consumed change in the same proportion as the quantities of goods produced, and deprives the formula of the property of reacting to changes in the technical input-output relations.

This leads one to suggest an index of the volume of net exports of the form $I_N = \Sigma(E_0V)/\Sigma(E_0)$, where E_0 stands for the value of net exports of any chosen subdivision of exports in the base year and V for a corresponding gross volume index. It is easy to verify that $I_N' = I_N$, if $q_1 = Vq_0$ for every export item within each subdivision of exports.

Basic Data

The Board of Trade index of the volume of exports takes 1947 as 100; it is therefore convenient to adopt this year as base of the net index, although 1948 would have served our purpose better.

In combining series for various subdivisions of exports to obtain aggregative index numbers each subdivision has been given a weight proportional to its estimated net exports in 1947. I have given elsewhere[5] details of the procedure adopted to estimate these weights, which are listed in Table I.

The Board of Trade has kindly allowed me to extract volume index numbers for the required subdivisions of exports. The net index is in effect a weighted arithmetic average of these gross indices. Table II presents the resulting net index numbers together with gross index numbers of the volume of exports and index numbers of the real product of the United Kingdom switched to 1947 = 100.[4]

Table II shows that the divergence between the gross and the net indices is small at first but that it tends to increase in 1950, 1951, and 1952. The net index is, except in 1948, consistently the higher of the two. Index numbers combining the first three classes of exports, that is food, raw materials, and manufactures, are also given; it is suggested that these should be used in any comparisons, since in the total index numbers (whether gross or net) the figure for parcel post is a mere guess while that for animals not for food is insignificant and can safely be omitted.

The divergence between the gross and the net indices is due to changes in the composition of exports. The greatest difference between them occurs in the food, drink, and tobacco class and indicates a considerable shift in this group towards commodities with low import content.

TABLE I

Weights for the index of the volume of net exports

Export Class and Group	Heading	Weight	
I	*Food, Drink and Tobacco*		43·4
A	Grain and flour	1·6	
B	Feeding stuffs for animals	0·1	
C	Animals, living, for food	†	
D	Meat	0·4	
E	Dairy produce	0·9	
F	Fresh fruit and vegetables	0·9	
G	Beverages and cocoa preparations (5)*	16·7	
H	Other food	14·9	
I	Tobacco	7·9	
II	*Raw Materials and Articles Mainly Unmanufactured*		32·3
A	Coal	2·5	
B	Other non-metalliferous mining and quarry products	2·0	
C	Iron ore and scrap	0·1	
D	Non-ferrous metalliferous ores and scrap	10·3	
E	Wood and timber	0·2	
F	Raw cotton, etc.	0·3	
G	Wool, raw, waste and rags (2)	6·3	
H	Silk, Raw, Waste and Artificial Silk Waste	1·3	
I	Other textile materials	0·4	
J	Seeds, nuts for oil, etc. (2)	1·2	
K	Hides and skins, undressed	1·2	
L	Papermaking materials	0·3	
M	Rubber	0·7	
N	Miscellaneous raw materials	5·5	
III	*Articles Wholly or Mainly Manufactured*		890·0
A	Coke and manufactured fuel	0·2	
B	Pottery, glass, abrasives, etc. (5)	29·1	
C	Iron and steel manufactures (2)	81·6	
D	Non-ferrous metals manufactures (5)	13·9	
E	Cutlery, hardware, etc.	33·1	
F	Electrical goods, etc. (3)	41·1	
G	Machinery (8)	173·0	
H	Manufactures of wood and timber	1·4	
I	Cotton yarns and manufactures (2)	61·2	
J	Woollen and worsted yarns and manufactures (6)	46·5	
K	Silk and artificial silk yarns, etc. (3)	27·3	
L	Manufactures of other textile materials	22·0	
M	Apparel (5)	25·2	
N	Footwear (2)	5·2	
O	Chemicals	63·0	
P	Oils, fats, etc. (3)	7·3	
Q	Leather and manufactures thereof	5·1	
R	Paper, cardboard, etc.	15·0	
S	Vehicles (4)	170·0	
T	Rubber manufactures	5·0	
U	Miscellaneous articles, etc.	63·8	
	Total Classes I, II, and III		965·7
IV	*Animals, not for Food*		4·3
V	*Parcel Post*		30·0
	Total All Classes		1 000·0
	Estimated Total of Net Exports in 1947		£ m. 967

* Since our formula assumes that the composition of exports within each export group is constant, some groups have been further subdivided; the number of additional subdivisions is given in brackets.

† Insignificant.

TABLE II

Index of the volume of net exports compared with the index of the volume of gross exports and the index of the real product of the UK, 1947 = 100

	1948	1949	1950	1951	1952
Total Exports					
Volume of gross exports	127·0	139·6	161·5	166·9	157*
Volume of net exports	126·9	140·5	162·2	170·7	159·6
Class I—Food, etc.					
Volume of gross exports	138·2	143·0	165·4	178·4	173*
Volume of net exports	142·0	148·6	168·2	187·3	179·9
Class II—Raw Materials					
Volume of gross exports	152·2	183·4	221·9	150·5	187*
Volume of net exports	160·3	187·4	227·4	157·0	185·7
Class III—Manufactures					
Volume of gross exports	126·0	138·9	160·9	168·4	155*
Volume of net exports	125·5	139·4	164·3	172·2	159·1
Classes I, II, III					
Volume of gross exports	127·5	140·5	163·1	168·5	157*
Volume of net exports	127·4	141·4	166·6	172·4	160·9
Real Product of the UK	105	111	114		
Real Product of the Manufacturing Industries	110	118	126		

* Linked back to 1947.

The difference between the indices in the raw materials class is explained by the increase in the coal exports. The net index is sensitive to changes in exports of commodities such as coal, since they have practically no import content.

The index numbers for manufactures show a much closer agreement. This result is produced by a number of conflicting changes, such as the increase in exports of metal goods, pottery, and vehicles, and a decrease in other textiles, apparel, and footwear. The increase in exports of cotton and wool manufactures influences the net index less than it does the gross index because of the high import content of these groups.

It is evident from the table that net exports have been growing much more quickly than the real product of the UK. Since the course of the index of the real product of the UK is flattened by the inclusion of public utilities and services, and since exports are mainly composed of products of the manufacturing industries, an index of the real product of manufacturing industries only is also given.

Conclusions

For many purposes the net index is preferable to the gross index as a tool for measuring changes in exports. In the period 1947–52 the divergence between the two indices is not large enough to affect any

economic conclusions arrived at by the use of gross indices. As, however, the two indices tend to diverge there is no reason why this situation should continue in the future.

The fact that the net index shows a greater increase than the gross index indicates that the export effort was successfully directed not merely towards increasing the total quantity exported but also towards the relative increase in 'more profitable' exports.

The comparison of changes in the real product of the UK with those of net exports shows that since 1948 an increasing proportion of the real product has been exported.

REFERENCES

[1] Board of Trade (1950). 'Volume Index Numbers of Imports and Exports, Change of Base Year,' *Board of Trade J.*, 24th June.

[2] Geary, R. C. (1944). 'The Concept of Net Volume of Output with special reference to Irish data,' *J. R. Statist. Soc.*, **107**, 251.

[3] United Nations Statistical Office (1950). 'Index Numbers of Industrial Production,' *Studies in Methods No.* 1.

[4] Carter, C. F. (1951). 'The Real Product of the UK, 1946–50,' London and Cambridge Economic Service *Bulletin*, **29**, 75.

[5] Billewicz, W. Z. (1952). *A Study of the Import Content of British Exports*, 1938 *and* 1946. Unpublished thesis, London University.

THE EVALUATION OF DETERGENT PERFORMANCE IN DOMESTIC DISHWASHING

CLIFFORD PUGH
The Shell Petroleum Co. Ltd., Thornton Research Centre

Statistical analysis is here used in the development of a practical dishwashing test for detergents. Mr Pugh gives a nomogram for assisting in testing the significance of results, and 'power curves' for showing what difference in performance can be significant for a given number of tests, or alternatively how many tests are required to make a given difference significant.

The widespread use of detergents for domestic dishwashing makes it desirable for manufacturers to carry out tests to evaluate the performance of their products in this field. As the housewife tends to use the foam of her wash liquor as a guide to efficiency, although, with the exception of the fatty acid soaps, foaming and detergency do not necessarily run in parallel, products intended for washing dishes must have a high standard of foam performance. Foaming is in consequence widely used as the main criterion of performance and is the basis of the test described here. Other tests are carried out on similar lines to investigate detergency.

Practical performance tests are usually difficult to control, as the absence of laboratory facilities makes precise control of the variables impracticable. To this must be added the difficulty of devising a comprehensive test procedure since there is great diversity in the way in which different housewives use detergents. It is clearly impossible to compare products under every set of conditions likely to be met in practice. It is also undesirable to adopt rigidly one set of test conditions and to adopt a specific, artificially prepared soil since it is then necessary to assume that the adopted procedure is representative of all conditions of use. One compromise between these two extremes is to permit considerable variation to occur in those variables, such as type of soil and method of washing, for which it is impossible to choose one representative level, but to fix on one or two values of more specific variables such as temperature, concentration, etc., where the range of variation is more readily defined. By the use of appropriate statistical designs, comparisons of products can then be made with reasonable precision, and the admission of the wider scope of some test variables permits the conclusions to be applied with more confidence than if a rigidly defined procedure is adopted.

This paper describes the development of a practical dishwashing test and indicates how statistical techniques can assist in overcoming some of the limitations inherent in such tests.

The Test Procedure

Since foaming is regarded as the main criterion of performance, the measure adopted is the number of plates washed before the foam is reduced to a thin surface layer. The five main factors which may affect the number of plates washed by a given product are

(i) the concentration of detergent,
(ii) the temperature of the wash liquor,
(iii) the hardness of the water,
(iv) the type of soil on the plates, and
(v) the method of washing used by the operator.

These can be divided into two sets. Items (i), (ii), and (iii) vary within a comparatively narrow range and can be specified for particular types of problem. For routine tests either a fixed concentration of the finished product (0·2 per cent.) or of the active matter (0·05 per cent.) is used at a temperature of 45° C., which is near the limit of comfort for the hands, and in hard water. For particular projects the effects of varying the concentration and of using soft water are also studied. It is interesting to note that in the cases so far studied there has been a linear relationship between the number of plates washed and the concentration over the range 0·025–0·1 per cent. active matter. This procedure is mainly of use to the manufacturer in studying the effect of variations in base detergents and other materials. For other purposes it may be desirable to compare products on an equal cost basis or, perhaps, to determine the concentrations required to give equal performance. The procedure to be adopted depends largely on the context of the problem being studied.

Item (iv) can vary very widely from one household to another. The difficulty of standardising the soil is overcome by using the plates from a works canteen for the test and adopting a randomised block technique in which the plates from any one course form a block. In this way the variation in the menu from day to day provides a wide variety of types of soil which can be regarded as an adequate sample of those occurring in household use. Differences in method to be found among different people in dishwashing, item (v), are not likely to be large. However, the tests in each block are carried out by one operator, who is instructed to remove all the soil from each plate using a dish mop in his or her own manner. A panel of seven operators is used in a random order.

It is considered that by allowing these two factors, soil and technique of washing, to vary in their normal manner from block to block the final comparisons among products will be more truly representative of general practice than had one type of soil and technique been adopted and rigidly standardised.

One practical limitation is the number of plates available in any one block. This permits of only four or occasionally five tests to be completed. When more than five products are to be compared a

balanced incomplete block technique[1] is adopted with either three or four tests per block.

Statistical Analysis of Results

In applying the analysis of variance technique the assumption is made that the results can be represented by a series of additive quantities:

$$y = a + p + s + o + \epsilon$$

where y is the number of plates washed,

a is a constant,

p is a component due to the product,

s is a component due to the soil,

o is a component due to the operator, and

ϵ is the normally distributed random error.

However, an inspection of the test results shows that they do not conform to this additive structure. The standard error tends to increase with the size of the reading, as is shown in Fig. 1, showing that the error component is multiplicative.

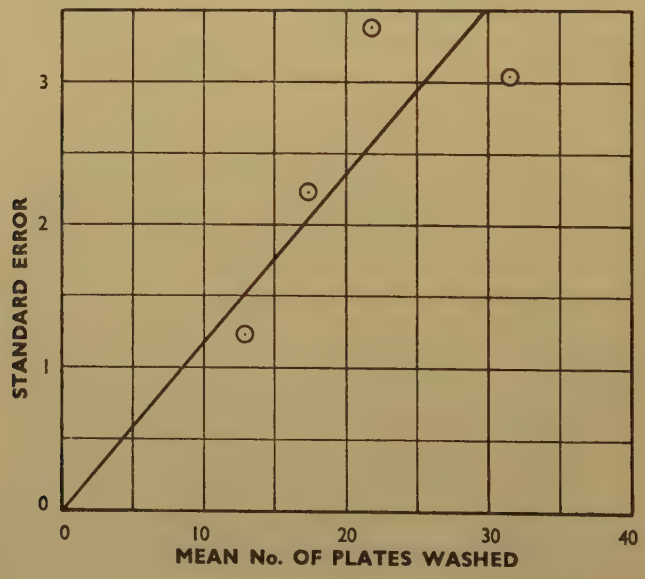


Fig. 1. Relationship between standard error and mean.

The relationship between products and soils is also found to be multiplicative. In Fig. 2 the numbers of plates washed by two products are plotted, each point representing a separate experiment, and hence a different soil. If the effects were additive the difference between the products would be constant and the figure would show a straight line

at 45 degrees to the axes with an intercept corresponding to the difference. If, on the other hand, the effects were multiplicative the figure would show a straight line passing through the origin with slope equal to the constant ratio of the performance of the products. Clearly the figure supports the latter hypothesis. The analysis of variance given in Table I confirms the multiplicative character of these results, and a similar examination of the results of other experiments has always led to the same conclusion.

In these circumstances the results can be represented by the relation $y = apso\epsilon$ and if the results are transformed to their logarithms,

$$\log y = \log a + \log p + \log s + \log o + \log \epsilon$$

and the assumption of additivity is justified. In addition, for the tests of significance to be valid the standard error must be independent of the mean; a logarithmic transformation also satisfies this requirement.

To obtain an estimate of the variance of the test, the results of twelve experiments have been analysed after transformation to logarithms. The estimates of variance attributable to experimental error are given in Table II. These are found to be homogeneous in each column

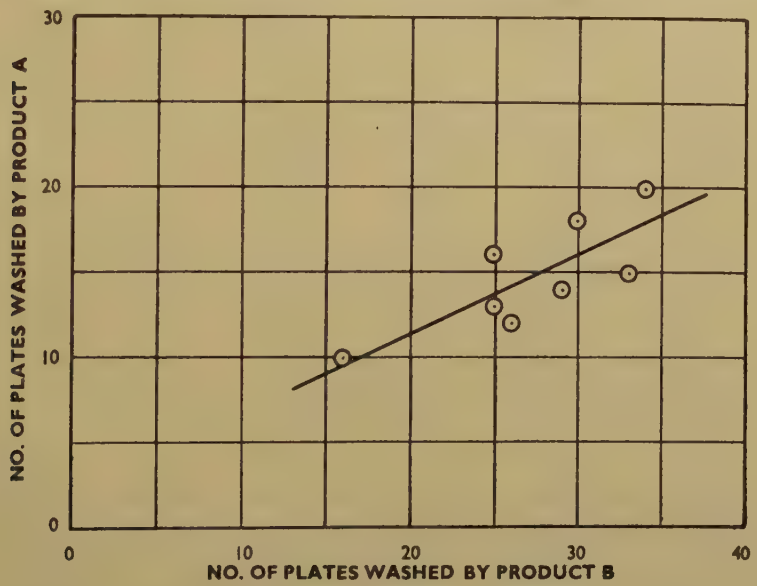


FIG. 2. Relationship between the performances of two products A and B.

when tested by the Bartlett test. The mean variances of the two types of experiment do not differ significantly (double-tail F-test) and the test variance is estimated at 0·0041 with 242 degrees of freedom.

Tests of Significance

When the analysis of results is carried out on the logarithms of the numbers of plates washed the tests of significance may be applied to

TABLE I

Analysis of variance of data represented in Fig. 2

Source of Variation	Sum of Squares	D.F.	Mean Square
Due to regression of result for product A on result for product B in same experiment	41·9	1	41·49*
Residual variation	32·01	6	5·335
Total	73·50	7	
Increase in residual variation caused by fitting a line through the origin (i.e. A/B constant)	1·79	1	1·79N.S.
Increase in residual variation caused by fitting a line at 45° to axis (i.e. A−B constant)	61·49	1	61·49*

* Significant at the 5 per cent. level.
N.S. = not significant.

TABLE II

Variances found in canteen dishwashing tests

Randomised Block Experiments		Incomplete Block Experiments	
Variance	D.F.	Variance	D.F.
		0·0092	11
0·0035	15	0·0077	11
0·0068	8	0·0048	11
0·0022	15	0·0034	35
0·0039	21	0·0037	35
0·0025	30	0·0044	35
		0·0044	15
Mean 0·0034	(89)	Mean 0·0046	(153)
Pooled variance		0·0041	(242)

the means of the transformed results in the normal way. It is more convenient, however, to test the significance of the results in the untransformed state once a sufficiently reliable estimate of experimental error has been obtained. This is done by testing the ratio of the geometric means by the relationship, derived from Student's *t*-test:

$$P_1/P_2\ (=r) > \text{antilog}\ tS_0\sqrt{2}/\sqrt{n}\ (=r_0)$$

where P_1 and P_2 are the geometric means of the results of n tests on each product,

S_0 is the estimated standard error of the test in logarithms, and

t is Student's t for the 5 per cent. level of significance.

When more than two products have been tested the Studentised range

test is substituted for Student's *t*-test since there is, in these circumstances, a greater probability that the largest ratio found will exceed r_0 than for any two products at random.

To enable the test to be completed as a routine by laboratory assistants these tests have been compiled in the form of a nomogram (Fig. 3), thus avoiding the necessity of consulting tables for *t* or the Studentised range test and calculating r_0 for individual experiments.

The rules for using the nomogram are simple:

(*a*) Place a straight edge to cut scale A at the determined variance (the value for these tests is marked with an arrow) and scale B at the number of washes carried out with each product. The edge then cuts scale C at the value of r_0 which must be exceeded to show a difference between two products significant at the 5 per cent. level.

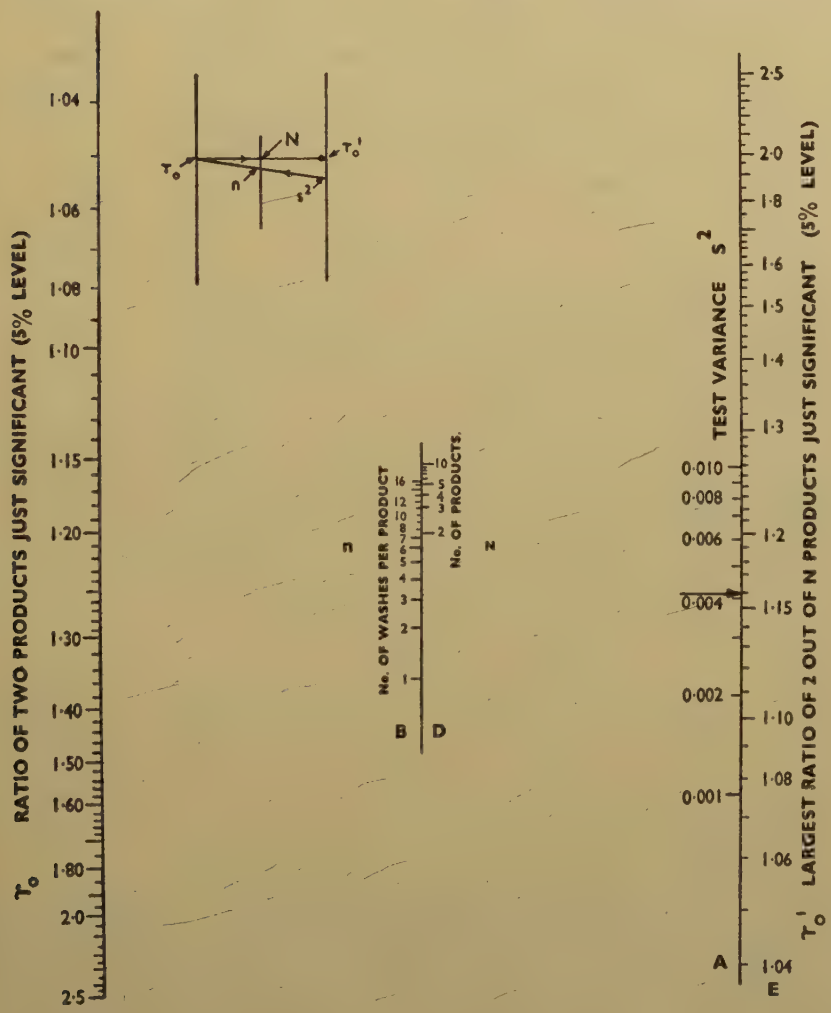


FIG. 3. Nomogram for testing the significance of results of the canteen dishwashing test.

(*b*) If the ratio to be tested is the largest among N products place the straight edge to cut scale C at the value r_0 determined and scale D at the value of N. The edge then cuts scale E at the value of r_0' which must be exceeded to show a difference between the products significant at the 5 per cent. level.

(*c*) If the ratio found is close to the ratio r_0 (or r_0') it is advisable to carry out further washes to confirm the result (or otherwise) before reporting.

(*d*) As a first approximation the arithmetic mean may be used, the geometric mean only being calculated when the ratio is on the borderline of significance.

In addition to testing the significance of an observed difference it is useful to know the probability of measuring as significant a ratio of any given size between two products in any given number of tests or conversely how many tests are desirable to give a reasonable probability of measuring as significant a given ratio if it exists. This probability, β, can be estimated from curves given by Pearson and Hartley[2] for the comparison of two products, using their curves for $\nu_1 = 1, \nu_2 = \infty$ (242 D.F.) to obtain values of a function

$$\phi = \frac{\frac{1}{2}\Delta}{S/\sqrt{n}}$$

where Δ in this case is the logarithm of the ratio concerned. Curves of β for various values of n the number of tests, and the ratio r representing the actual ratio of the efficiencies of the products, are given in Fig. 4, where β is the probability of the result of n tests being significant at the 5 per cent. level.

As illustration Table III gives the number of washes required if it is necessary to detect a given difference as significant, at the 5 per cent. level, in nine out of ten experiments. It will be noticed that to detect an improvement of 25 per cent. a minimum of ten replications is required. This high number is typical of practical tests, a point which is frequently overlooked in designing experiments.

TABLE III

Real Ratio of Products	1·25	1·33	1·5
No. of Tests Required for $\beta > 0{\cdot}9$	10	6	3

Conversely, Fig. 4 may be used to determine how many tests are needed to confirm that the actual difference between two products does not exceed a given value. Thus, to show that the difference between two products is less than 20 per cent. (with a 1 in 10 risk of being wrong) the ratio of the means of fourteen tests on each product should not be significant.

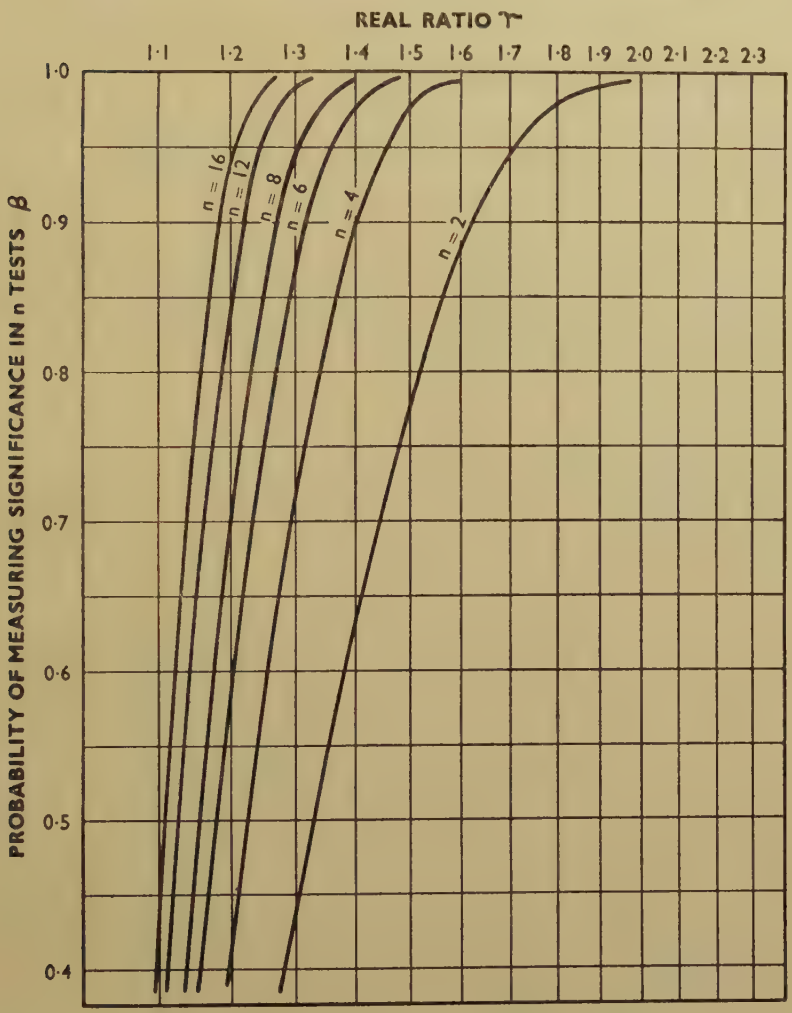


FIG. 4. Power curves of the canteen dishwashing test.

The author thanks the directors of the Shell Petroleum Co. Ltd for permission to publish this paper and the many colleagues who have assisted in the development and carrying out of these tests.

REFERENCES

[1] BROWNLEE, K. A. (1949). *Industrial Experimentation*, p. 149, 4th edition. HMSO, London.

[2] PEARSON, E. S., and HARTLEY, H. O. (1951). 'Charts of the Power Function for Analysis of Variance Tests derived from the Non-central *F* Distribution,' *Biometrika*, **41**, 112.

THE USE OF SAMPLES IN AUDITING

NOEL WILLIAMS, FIA

Mr Williams shows how the ordinary statistical theory of sampling provides the auditor with criteria for deciding from checks on samples of items whether the whole set of items can be accepted as correct. Readers may care to comment on the procedures suggested.

As one part of his check, an auditor will test the accuracy of a number of entries in the various sections of the accounts of a company. He may be able to check all the entries in a few of the sections, but usually the number of transactions is so great that he can check only a sample. For example, the number of wage or salary payments in any week or month may be several hundreds or even thousands. Before checking the deductions and net payments the auditor will wish to satisfy himself that the gross wages or salaries and their total, which is the charge to the company, are correct. To be sure that every item is correct he must check each one and their addition. If he checks only a sample of the items and finds no errors, is he taking a reasonable risk in accepting the others and their total as correct? The errors, if any, may be clerical or may be due to fraud. If the auditor finds any serious errors in his sample items he will have to make a more extensive check, and sampling methods will not then be appropriate. The object of this paper is to suggest how some well-known statistical theorems can be applied to assist the auditor to make a decision.

The following notation has been used:

$\mathcal{N}$ denotes the total number of entries in a given period in a specified section of the accounts, for example the number of wage payments in a given week, termed the 'population' of entries,

n denotes the number of entries to be examined, and

x denotes the value of an entry.

If we sample with replacement, i.e. return each item to the whole before taking another for the sample, the chance of drawing n values all of which are correct from a population of $\mathcal{N}$ containing $q\mathcal{N}$ errors is $(1 - q)^n$. If $\mathcal{N}$ is large and n and $q\mathcal{N}$ are both small, and we sample without replacement, as we usually do in practice, the corresponding chance is approximately $(1 - q)^n$. Until we have examined the items we do not know the value of q, which may of course be zero; we must therefore specify the highest proportion of errors that would be acceptable, say q'. We must also take into account that, although the population may have at least the maximum permissible proportion of

errors, we may draw a sample consisting of values which are all correct, in which case we shall not know that there are any errors. For example, suppose that $q' = 0{\cdot}01$ and that we have drawn a sample of, say, 100 items and on checking them found no errors; then the chance of making a wrong decision, i.e. accepting a population of entries containing more than 1 per cent. of errors, is $(1 - 0{\cdot}01)^{100} = 0{\cdot}37$ and if this is too great we must increase the size of our sample. If the chance of making a wrong decision of this kind is required to be as low as 0·05 our sample must consist of 296 items, all correct.

The above method of checking a sample of sufficient size and, if all items are correct, accepting the population as correct, may be suitable in some cases—for example when checking that requisitions have been signed by an authorised official, where the item can be only right or wrong. It is, however, not so suitable when checking the *value* of an entry because no account is taken of the size of any errors, so that although the number of items which are wrong may be small, the value of these errors may be important. A similar objection applies to any scheme of sampling based on classifying items as correct or wrong. It was this objection which led to the application of the method described below.

Let T_0 denote the total value of the N entries, as shown by the books of account; then the average value is T_0/N. We draw a sample of n items and check them, and if they are all correct find their average value $\bar{x}$ and their standard deviation σ_x. Let $N\bar{x} = T_1$, so that T_1 is an estimate of the total value of the entries, based on values which have been checked.

It would, of course, be possible at this stage to compare the sample mean, $\bar{x}$, with T_0/N to see whether the difference could be accounted for by sampling fluctuations. It is, however, more convincing from an auditor's point of view to test T_0 against T_1 to see whether T_0 can be accepted; in other words, can the difference between T_0 and T_1 be accounted for by sampling fluctuations, or does it suggest that T_0 may be incorrect? Even though we cannot be certain that a small difference is due to sampling fluctuations we may be able to say that T_0 is probably correct or at worst is not likely to be seriously wrong.

The estimate T_1 has a percentage standard error of $100\sigma_x/\bar{x}\sqrt{n}$, and this provides a criterion with which to compare $100(T_0 \sim T_1)/T_1$, i.e. the percentage difference between our actual and estimated totals. If our sample is drawn from a normal population we could accept a percentage difference between T_1 and T_0 up to twice, or perhaps three times, the percentage standard error of T_1. In practice we are sampling from populations which are often not even approximately normal, and caution must therefore be exercised in setting limits to the acceptable percentage difference. A more important restriction on the acceptable deviation of T_1 from T_0 is the size of error which the auditor would regard as serious. For example, suppose that a difference of 4 per cent. is regarded as serious; then if the difference between the sample estimate

T_1 and T_0 is 4 per cent. or more, it would be unwise to accept the population even though the difference could be accounted for by sampling fluctuations. For practical work it appears desirable to restrict acceptance to cases where the percentage standard deviation is small and the percentage difference does not exceed it. If these conditions are not fulfilled, a further sample should be drawn. The limits must be restricted because if some of the unchecked values are incorrect, and hence T_0 does not represent the true total of the correct values, then we may draw a sample of correct entries which, because of sampling fluctuations, may happen to give an estimate close to the incorrect T_0. Two examples follow.

Example I. The data were monthly gross salaries of 232 employees, totalling £11 749 10s. 1d. $= T_0$. A sample of 46 was drawn and all the items were checked. The total value was £2 309 10s. 1d. Then $T_1 =$ £2 309 10s. 1d. $\times 232/46 =$ £11 647·933 and $100(T_0 - T_1)/T_1 = 0{\cdot}87$ per cent. The distribution is in Table I. From the grouped data of the sample the standard deviation was found to be 11·17 and the percentage standard deviation of T_1 to be 3·3. It will be seen that the percentage difference between the actual and estimated totals is less than the percentage standard deviation of the estimate, and we may accept T_0 as being substantially correct.

TABLE I

Gross Salary (nearest £1)	Sample Frequency
27–29	1
30–32	2
33–35	5
36–38	1
39–41	4
42–44	4
45–47	3
48–50	—
51–53	1
54–56	7
57–59	5
60–62	11
63–65	1
66–68	1
Total ..	46

Example II. The data were monthly gross salaries of 367 employees, totalling £11 841 16s. In this case all salaries under £20 and over £45 were extracted and checked; this reduced the variation in the remaining values and also set a limit on the error that could be contained in any one value. This reduced the number of entries to 354 $= N$, and the total value to £11 491 12s. 3d. $= T_0$. A sample was drawn and checked, but as the results were not considered to be close enough a further sample was taken. No errors were found. The results are given in Table II.

TABLE II

Sample		A	B	Combined
Gross Salary		Sample Frequency		
Not less than £	Less than £			
20	23	3	2	5
23	26	6	6	12
26	29	1	5	6
29	32	6	1	7
32	35	4	3	7
35	38	5	12	17
38	41	3	3	6
41	44	8	3	11
Number of items = n		36	35	71
Sample total salary		1 195·533	1 155·333	2 350·866
Sample mean = $\bar{x}$		33·209 26	33·009 52	33·110 80
Estimate = T_1		11 756·08	11 685·37	11 721·22
Standard deviation = σ_x		7·4	6·4	6·9
Percentage difference = $100(T_0 \sim T_1)/T_1$		2·2	1·7	2·0
Percentage standard deviation of T_1		3·2	3·2	2·3

By combining the samples the percentage standard deviation of T_1 was reduced to 2·3 and as the actual percentage difference of 2·0 was less than this, T_0 was accepted as correct.

The samples in the above two examples were drawn systematically by choosing a number up to 5 or 10 at random, drawing the corresponding item as the first in the sample, and thereafter drawing every fifth item in Example I and every tenth in Example II. This is more satisfactory than drawing a purely random sample, as the items may cover several subdivisions; for example, the salaries covered several departments and the auditor will wish to be sure of checking a few from each subdivision. Provided that care is taken to avoid any recurring features in the items in the population, this method of drawing the samples introduces some stratification and will tend to reduce sampling errors.

If it is considered desirable that the percentage standard error of T_1 should not be greater than some value C, then our sample size will be found by solving the equation $C = 100\sigma_x/\bar{x}\sqrt{n}$, from which the sample size is

$$n = (100\sigma_x/C\bar{x})^2$$

It is, however, necessary to start with a sample of arbitrary size from which the standard deviation can be estimated.

PLANNING AN EXPERIMENT IN A COTTON SPINNING MILL

ROBERT E. PEAKE

British Cotton Industry Research Association

Mr Peake writes: 'There are not very many published examples of planned experiments conducted in factories, and even fewer published accounts analyse the technical and practical considerations that lie behind the choice of the particular plan adopted.' This article is a contribution towards filling the gap.

Introduction

It is a commonplace that a well-planned experiment leads to more reliable results more economically than an experiment arranged without full consideration of the variables and errors involved, and statistical literature abounds in examples of planned experiments, in standard experimental designs, and in expositions of how to deal with the results. When one attempts to plan an experiment in a particular situation, the facts of the situation prove to be as important as statistical principles, and some experience in the field of application is necessary for success. There are not very many published examples of planned experiments conducted in factories, and even fewer published accounts analyse the technical and practical considerations that lie behind the choice of the particular plan adopted. This article describes an experiment conducted in a cotton spinning mill, and particular attention is paid to the genesis and development of the investigation.

Experimental Conditions and Sources of Error

The experiment concerned one of the preparatory processes of spinning, the production of roving, which is a strand of cotton thicker than the final yarn or thread. Roving is produced on a roving frame, which has about 180 spindles each producing a bobbin of roving; associated with each spindle is a rotary guide, termed a 'flyer,' which by rotation puts twist into the strand and at the same time guides it on to the bobbin. Plate I shows a roving frame with the operative attending to the cause of a stop. The twist is inserted to give the roving sufficient cohesion to withstand the tensions of processing, both at this stage and subsequently. The degree of cohesion increases with the degree of twist (measured by the turns per inch), but an increase in twist reduces the rate of production and so adds to the cost of processing. Recently some special flyers have been devised to insert a 'false twist' at a place where the roving is specially vulnerable, the twist being false in the sense of being localised and temporary and having no effect on the turns per inch in the final roving. Part of the

purpose of the experiment was to investigate the effect of varying the twist for each type of flyer and part was to compare the ordinary and special flyers.

The turns per inch commonly used for the roving under investigation are 1·69 and 1·78 with the special and ordinary flyers respectively, and preliminary tests showed that the turns could be reduced to 1·63 for special but not for ordinary flyers and increased to 1·90 for the ordinary flyers. Accordingly the six experimental conditions shown in Table I were chosen for full investigation. They are labelled A to F. It is perhaps an interesting practical point of this experiment that before designing the experiment it was necessary to make these preliminary tests to decide the range of twist to be covered.

TABLE I

Experimental conditions

Type of flyer	Twist (turns per inch)			
Ordinary ..	—	1·69(A)	1·78(B)	1·90(C)
Special ..	1·63(D)	1·69(E)	1·78(F)	—

From time to time a strand of roving breaks during processing and the frame has to be stopped to mend it. An operative is in charge of several roving frames and spends a large part of the time in mending such breaks; moreover the quality of the final product is also to some extent dependent on the number of places that have been mended. In fact, the breakage rate, expressed as the number of breaks per 100 pounds of material, was chosen as a realistic criterion of the effects of twists and flyers on the processing efficiency and quality of the roving. The experimental unit was taken as the production of one set of full bobbins on a complete frame. It had to be a full set because the breakage rate may undergo a trend during the machine cycle from empty to full bobbin, and it had to be a complete frame because of the technical impossibility of using different twists or different flyers simultaneously on the same frame.

As these tests were to be made under factory conditions and, even if we observed several roving frames, the experiment would take several weeks, we expected variations in breakage rates additional to those due to the experimental conditions. The two main potential sources of this unwanted variation would be (1) differences between machines, accompanied by differences between the skills of the operatives in charge of the machines, and (2) changes with time in the material supplied to the machines and in atmospheric conditions. Nominally similar machines and material would be used all the time and a plant controlling the relative humidity in the factory would be in operation, but nevertheless some variation was to be expected from these sources. Accordingly it was desirable to design the experiment

to cover these effects and essential to make sufficient observations to reduce the residual error from these sources to a suitable size.

Experimental Designs

The first plan to be considered was to use six machines and to dispose the six experimental conditions upon them in the form of a Latin square, shown in Table II. Each experimental condition occurs once, and once only, in every row and in every column, and hence the two sources of error are equally spread over all conditions.

TABLE II

Design 1

Machine \ Time	T_1	T_2	T_3	T_4	T_5	T_6
M_1	F	E	B	A	C	D
M_2	A	F	D	C	B	E
M_3	D	C	E	F	A	B
M_4	B	A	C	D	E	F
M_5	C	B	F	E	D	A
M_6	E	D	A	B	F	C

This design controls the effect of both the main sources of unwanted variation but it has the drawback of requiring six machines. Under the conditions prevailing at the factory the number of machines available was liable to fluctuate, and it was even unlikely that six would ever be simultaneously available. In the event only two machines were available. We also considered the use of quasi-Latin squares, which permit the use of fewer machines than experimental conditions. We rejected these, however, because they require that the number of machines is unchanged during the experiment.

Accordingly a randomised block design was next considered, which allowed control of the unwanted variation with fewer than six machines and permitted the number of machines to vary during the experiment. This is shown in Table III.

TABLE III

Design 2

Order of Testing \ Block	1	2	3	4 . . ., etc.
1	E	D	F	F . . .
2	A	C	D	D . . .
3	C	A	C	A . . .
4	D	B	A	. . . .
5	B	E	B	. . . .
6	F	F	E	. . . .

In design 2 the six experimental conditions are observed randomly one after the other on one machine to form a 'block.' The next block

may consist of another random sequence made on a second machine at the same time as when the first is done, or it may follow block 1 on the same machine. Changes that occur with time during a block are to be averaged by accumulating a large number of blocks, and differences between machines and operatives and long-term time changes are controlled by the fact that each machine and operative and block is used for all six conditions.

Both designs 1 and 2 have the drawback that frequent changes of the experimental conditions must be made and, although a change of twist is fairly easy to accomplish, the changing of the flyers is a relatively long operation and so might hamper the factory production. To reduce this interference with production a third design was considered in which one half of a block is confined to conditions ABC and the other half to DEF, thereby reducing the number of stoppages for flyer changing. The type of flyer to be used in the first half of the block is decided randomly and the three twists associated with that type of flyer are used in random succession before the flyers are changed. The third design is shown in Table IV.

TABLE IV

Design 3

Order of Testing \ Block	1	2	3	4 . . ., etc.
1	E	A	A	F . . .
2	D	C	B	E . . .
3	F	B	C	D . . .
4	B	D	E	B . . .
5	A	E	F	C . . .
6	C	F	D	A . . .

This design has the disadvantage that the comparisons of flyers are separated by longer intervals of time than those of twists, so that the twist comparisons gain in accuracy at the expense of the flyer comparisons. Now, on an average, a flyer change occurs one and half times per block in design 3 as against three times per block in design 2, and this represents a gain in productive time of 45 minutes per block. As a block would take about a week to complete it was decided that the saving was not worth while at the expense of statistical efficiency, and design 2 was selected as the final design for the experiment.

The Size of the Experiment

Having decided on design 2 we wished to decide how big the experiment should be, judging this by the number of blocks to be taken to reduce the residual error to a suitable size. The size is suitable only if it permits the identification of the smallest difference between the

results of any two experimental conditions that is of practical importance to the purpose of the investigation. Occasionally, the purpose defines this smallest difference completely; more often, as here, a rough value has to be assumed more or less arbitrarily. In this particular experiment we expected the average breakage rate to be roughly 10 breaks per 100 pounds and thought that only differences of less than 20 per cent. would be unimportant. Accordingly we made our calculations on the basis of detecting a difference of at least two breaks per 100 pounds with reasonable certainty. The exact determination of the number of blocks involves theory beyond the scope of this account, but in view of the arbitrary nature of the criterion a very approximate method was used.

To avoid inconvenience to the mill staff, two sets of bobbins, rather than the ideal of one, were observed after each change of the experimental conditions. Now, if two sets of bobbins are observed under each experimental condition, n blocks are taken, and the error variance of one set is σ^2, the error variance of the average for each experimental condition is $\sigma^2/2n$, and that of the difference between any two such averages is σ^2/n. Then, on the basis of a 95 per cent. confidence level, if the true values for two experimental conditions are the same, the differences between observed averages might be as high as $2\sqrt{(\sigma^2/n)}$. If, however, there was a true difference of 2, the lowest value to be expected for this would be $[2 - 2\sqrt{(\sigma^2/n)}]$. Hence we must choose n so that

$$[2 - 2\sqrt{(\sigma^2/n)}] \geqslant 2\sqrt{(\sigma^2/n)}$$

that is

$$n \geqslant 4\sigma^2$$

As the precise value of σ^2 was unknown we used the best value available —one obtained from a previous experiment of quite a different character—and this suggested that n should be 56. The fact that σ^2 has been so estimated makes the reasoning just given subject to some theoretical objections, which for our purposes were unimportant. Readers who require a more rigorous approach to the problem of determining the size of an experiment are referred to Chapter 2 of *Experimental Designs* by Cochran and Cox,[1] to the references at the end of that chapter, and to a paper by Sillito.[2]

A block was expected to take about a week to complete, and with only a few machines available 56 blocks would have taken a longer time than could be lightly afforded. Accordingly we decided to examine the results after about one-quarter of the 56 blocks had been completed and to take a further decision in the light of the information thus available. In fact we examined the data for the first 13 blocks using, for ease, the usual analysis appropriate to an experiment of predetermined size instead of a sounder 'sequential' analysis. Although the adopted procedure is open to theoretical objections, our practical conclusions were established with a very high degree of statistical significance and these objections do not invalidate them.

PLATE I. Part of a roving frame. The material on bobbins at the top is coarser than roving and is being fed into the machine. The roving is being wound on to the bobbins towards the bottom of the picture. The false twist is inserted in the region indicated by the white arrow. The operative is attending to a cause of frame stoppage. (Photograph by courtesy of the Cotton Board.)

The Results of the Experiment

The average breakage rate for each experimental condition is given in Table V and in Fig. 1 is plotted against turns per inch. In Table VI (*a*) the variance of these averages is analysed into the three parts appropriate to a randomised block experiment.

The breaks observed in processing roving are an accidental phenomenon, and the error variance is proportional to the mean breakage rate. In order to satisfy the assumption of homogeneity of variance lying behind the usual method of analysis it would have been better to treat the square root of the results in Table V. Against this, the square root of the breakage rate has less practical meaning, and as this consideration weighed heavily no transformation of the data was made. We believe that the practical conclusions are nevertheless valid.

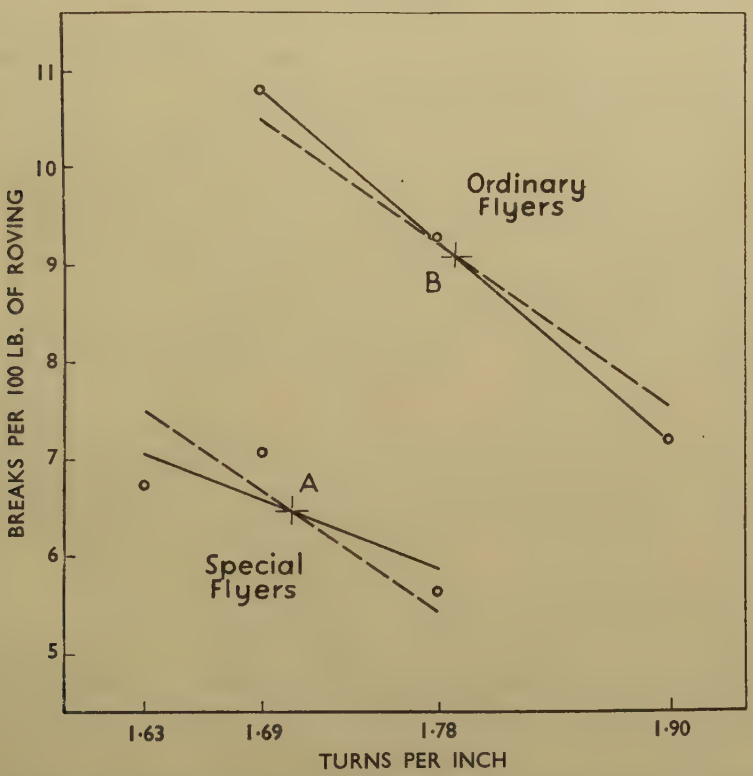


FIG. 1. Relationship between breakage rate and twist for two types of flyer. Each full regression line has a different slope appropriate to the flyer; the broken lines have the same slope.

It is evident from Table VI (*a*) that not only did the breakage rates for different experimental conditions differ, but so did the rates for different blocks; this emphasises the value of having designed the experiment to remove such effects from the vital comparisons. Further, the confidence interval on the difference between the averages for any two experimental conditions is $2\sqrt{(10{\cdot}2/13)}$, that is 1·8, and thus we

TABLE V

Experimental results, breaks per 100 *pounds of material**

Experimental Condition	Twist (t.p.i.)	Mean for Experimental Condition	Block No. 1	2	3	4	5	6	7	8	9	10	11	12	13
A	1·69	10·8	6·0	9·7	7·4	11·5	17·9	11·9	10·2	7·8	10·6	17·5	10·6	10·6	8·7
B	1·78	9·3	6·4	8·3	7·9	8·8	10·1	11·5	8·7	9·7	8·3	9·2	9·2	10·1	12·4
C	1·90	7·2	2·3	3·3	7·3	10·6	7·9	5·5	7·8	5·0	7·8	6·4	8·3	9·2	12·0
D	1·63	6·8	3·3	6·4	4·1	6·9	6·0	7·4	6·0	7·3	7·8	7·4	7·3	10·1	7·8
E	1·69	7·1	3·7	6·4	8·3	3·3	7·8	5·9	8·3	5·1	6·0	3·7	11·5	13·8	8·3
F	1·78	5·7	4·2	4·6	5·0	4·1	5·5	3·2	10·1	4·2	5·1	4·6	11·5	5·0	6·4

* Each block result is the average of two unit tests. t.p.i. signifies turns per inch.

TABLE VI

Analysis of variance of breakage rate data

Source of Variation	D.F.	Mean Square	Ratio[1]	Significance[2]
(*a*) *Randomised blocks*				
Between experimental conditions	5	92·4	9·1	***
Between blocks	12	29·5	2·9	**
Residual error	60	10·2		
(*b*) *Experimental effects*				
Deviations of treatment means from separate regression lines	2	5·1	0·5	—
Differences in slope between two regression lines	1	16·7	1·6	—
Flyer and twist main effects	2	217·5	21·3	***
Total	5	92·4	9·1	***
(*c*) *Partition I of main effects*				
Flyer effect	1	426·5	41·8	***
Remainder I	1	8·5		not tested
Total	2	217·5	21·3	***
(*d*) *Partition II of main effects*				
Twist effect	1	173·7	17·0	***
Remainder II	1	261·3		not tested
Total	2	217·5	21·3	***

[1] Ratio of mean square to error mean square.
[2] — not significant, ** significant between the 1 and 0·1 per cent. levels, *** significant above the 0·1 per cent. level.

see that the special flyers gave smaller breakage rates than did the ordinary flyers when the same twist was used at each of the twists 1·69 and 1·78. Moreover, the special flyers with the lowest twist gave a smaller breakage rate than the ordinary flyers with their normal twist.

It is, however, advisable to extend the analysis of the differences between results for the experimental conditions, using regression analysis to study the effect of twist.

A first analysis of the effects of the experimental conditions is in Table VI (*b*). The first row shows that the variance of the six averages about the individual regressions on twist for each flyer (the full lines of Fig. 1) is not significantly different from the error variance of 10·2 and hence there is no objection, implicit in the data, to the assumption of a linear relation between breakage rate and twist. The second row tests whether these full lines really differ in slope, that is whether there is an interaction between the flyers and twist, or whether one slope of regression on twist, shown in the figure by the parallel dotted lines, is sufficient to describe the data. The variance due to differences in slope is not significantly greater than the error variance. The third row represents jointly the effect of differences between the breakage

rates for the two flyers and the effect on the breakage rate of the common regression due to twist. This is very highly significant.

In section (*c*) of Table VI these flyer and twist main effects are analysed to test the flyer effect. This is done by testing a variance, given in the first row, which results from the differences in slope between the dotted lines and the line joining A and B, the two average rates for the flyers. In this way the distance parallel to the y-axis between the dotted lines is shown to be highly significant—most readers will agree that this is obvious from the diagram. The remainder I has no useful physical meaning.

In section (*d*) of Table VI the variance due to both main effects is again partitioned to test the significance of the relationship between breakage rate and twist. The first row in this section shows that the common slope of the dotted lines gives a highly significantly better fit to the data than would horizontal lines representing no relationship. An increased twist reduces the breakage rate. Again the remainder has no useful physical meaning.

Should we now have continued the experiment? By doing so we would have estimated the effect of flyer and twist with greater precision and have investigated more closely the possibilities of an interaction between twist and flyers and of departures from linearity of the relationship between breakage rate and twist. These matters were, however, of little technical interest and we decided to terminate the experiment as designed, being content with having established the existence of the main effects.

Any further experiments at this mill would be better directed towards more particular comparisons. For example, it might be useful to study the relationship between breakage rate and twist more closely for the special flyer alone, since it is now the behaviour of this flyer that is of primary interest.

REFERENCES

[1] Cochran, W. G., and Cox, G. M. (1950). *Experimental Designs*. Pp. 15 et seq., Wiley, New York, and Chapman & Hall, London.

[2] Sillito, G. P. (1948). 'The Numbers of Observations needed in Experiments leading to a *t*-Test of a Mean or the Difference of Two Means,' *Research*, 1, Supplement 11, p. 520.

A CHANGE IN THE PROPORTION OF JURORS

THOMAS CORLETT

The Social Survey

In an earlier article in *Applied Statistics* Mr Corlett described how the proportion of electors liable for jury service can be used as an index of the economic status of an area, and this index has been used in the design of samples. Recently there has been a change in the method of determining liability for jury service, but Mr Corlett shows here that the change makes little difference to the index for the uses originally proposed.

In an article[1] in the first issue of this journal I described how the proportion of electors in any area in England and Wales who were liable for jury service, those marked J in the electoral register, could be used as an index of the relative economic status of the area. The chief reason for this relationship is that the main qualification for jury service is to be a householder living in premises of more than a certain rateable value.

This index is being used by the Social Survey as a way of stratifying small areas, such as wards, parishes, or polling districts, when these areas are used as sampling units in two- or three-stage sample designs for England and Wales. The idea has been taken up by a number of commercial organisations. But since this recently aroused interest of the sample-survey practitioner in the rather unlikely subject of jury service there has been a change in the method of determining the liability of people to serve on juries and hence a change in the proportion of people liable. The purpose of this article is to draw the attention of users of the index to this change and assess its size and possible effect on the efficiency of the index.

A Social Survey report[2] referred to in the original article included a number of analyses of a small sample of 201 people marked as being liable for jury service in the 1951 Register of Electors. One was an age-analysis which showed that 54, or 27 per cent., of jurors were over 60 years of age and were thus actually unqualified for jury service. In the year following the publication of the report the Home Office took steps to eliminate these unqualified jurors.

In November of every year a standard form of return is issued to householders in order that the new register of electors, due to be published the following March, may be compiled. On this the householder has to enter the full names of all persons aged 21 or over in his household living on the premises on 20th November. The form issued in November 1952 was amended so that the name of the 'occupier' of the property would be distinguishable. (The 'occupier' is, of course, the person who is qualified and liable for jury service if the property

is above the rateable value limit and if he or she is not over 60.) The occupier was also asked to state if he was over 60, and a note was added to the form explaining the reason for this extra question.

We had originally estimated that the number of jurors on the 1951 register was about 1 600 000, some 5·3 per cent. of the total electorate. Clearly, if 27 per cent. of these were over 60 and all were identified as such by means of the revised form of return issued in November 1952, the 1953 register would be expected to contain 27 per cent. fewer jurors, other things being equal. Soon after copies of the 1953 register were available we made arrangements to check the actual reduction in numbers.

A sample of 54 polling districts, selected with probability proportional to their 1952 electorates, was taken: these were in fact the second-stage units in a three-stage sample design for England and Wales. For each of these polling districts the electorate and the number of jurors in the 1951, 1952, and 1953 registers were counted. On the basis of this sample the estimates given in Table I were made. The estimated reduction in the number of jurors from 1952 to 1953 was thus 13 per cent. It may perhaps be assumed that, if no special action had been taken, the number would have slightly *increased* during the period comparably with the slight increase from 1951 to 1952. On this assumption the reduction in numbers attributable to the special action would be about 15 per cent. This latter figure is confirmed by the

TABLE I

Estimates for England and Wales

Year of Publication of Register	Total Electorate* *millions*	Percentage who are Jurors	Total Jurors *millions*
1951	30·74	5·40	1·66
1952	30·81	5·50	1·70
1953	30·77	4·78	1·47

* The figure for 1951 is the total of local government electors given in the Registrar General's Statistical Review of England and Wales, 1951. Those for 1952 and 1953 are estimates based on the sample of polling districts.

value of the regression coefficient in the equation given below. This was derived from fitting a straight line to the values of the proportion of the electorate who were jurors ($\mathcal{J}$) in each of the 54 polling districts in 1952 and 1953 respectively.

$$\mathcal{J}_{53} = 0{\cdot}13 + 0{\cdot}85\mathcal{J}_{52}$$

In order to reconcile the 'expected' reduction of 27 per cent. with the actual reduction of 15 per cent., the 54 over-age jurors in the original sample from the 1951 register were looked up in the 1953 register to see what proportion of them had actually ceased to be marked J. Ten

of them were no longer registered at the same address, but of the 44 who were still there 18 were still marked J and 26 were no longer so marked. On this basis a more realistic 'expectation' of the reduction due to the change in the form of return would be roughly $26/44 \times 27$, or about 16 per cent., which agrees closely with the independent estimate of the actual reduction.

In considering the changes in the numbers of jurors from year to year it is of interest to see how the changes are likely to affect the efficiency of the J-index itself. The correlation coefficients between the values of the index for the different years in the sample of polling districts referred to above are as follows:

$$r_{51/52} = 0{\cdot}997$$
$$r_{51/53} = 0{\cdot}971$$
$$r_{52/53} = 0{\cdot}974$$

From the high value of all these coefficients it may be inferred that the changes themselves can have had little effect on the efficiency of the index as a method of ranking areas by economic status. Values of the index derived from the 1951 register, for example, can still be used today; there would be little, if any, advantage in using 1953 values in their place, though it would, of course, be wrong to combine values from two different years in a single series.

REFERENCES

[1] Corlett, T. (1952). 'A Use for the Jury Qualification in Sample Design,' *Applied Statistics*, **1**, 34.

[2] Gray, P. G., Corlett, T., and Jones, Pamela. (1951). *The Proportion of Jurors as an Index of the Economic Status of a District.* London: The Social Survey.

NOTES AND COMMENTS

The Institute of Statistics of the University of Paris has established 'Le Centre de Formation des Ingénieurs et Cadres aux Applications Industrielles de la Statistique' for the purpose of fostering the application of statistics to industry in France. The Centre is arranging two types of courses of instruction, one an elementary course lasting for a period of between ten and fifteen days and designed to teach the principles and practice of statistical quality control, and the other for engineers with some mathematical training, lasting for three weeks and covering a considerable amount of statistical theory as well as sampling and market research. It would be useful to know what sorts of people successfully attend such short but compressed and comprehensive courses and what the students get out of the courses.

The organ of the Centre is the *Revue de Statistique Appliquée*, the first number of which has recently appeared. This journal, which is edited by M. E. Morice, aims at presenting concrete examples of the application of statistics to industry, taken from France and abroad, and expositions of statistical methods. The articles in the first number will undoubtedly further the aims of the journal and Centre. We wish this new enterprise every success.

★ ★ ★

Applied Statistics is now available to students attending courses in statistics at the reduced rate of 15s. per volume. Each application should be sent to the publishers together with a letter from the teacher or other authority stating that the student is attending the course. One volume of *Applied Statistics* will be supplied at the reduced rate for each session of attendance.

Publishers' binding cases in blue buckram, stamped in gold on the spine, are now available for each volume of *Applied Statistics* at a cost of 5s. 6d. post free. These may be ordered direct from the publishers. Subscribers may send their parts to the publishers for binding. The total cost of supplying the binding case and binding is 18s. for each volume, post free.

A limited number of most of the earlier parts are available to complete sets. Application for these may be made to the publishers, either directly or through the normal suppliers.

INDUSTRIAL APPLICATIONS SECTION OF THE ROYAL STATISTICAL SOCIETY

List and Reports of Meetings

BIRMINGHAM and DISTRICT GROUP

25th March 1953. A Problem in Ranking: M. G. Kendall

After describing some possible approaches to the problem of comparing rankings, Professor Kendall showed how the possible measures could be developed from the league system of scoring in which the winner of each game receives the same score irrespective of the quality of the opposing team. With this system as a base a new system of scores can be developed in which each winner is allotted a score proportional to the total score of the beaten opponent. After several repetitions of this process a stable ranking emerges.

29th April 1953. Statistical Quality Control and the Analysis of Variance: V. E. Gough

It is often difficult to interpret the physical meaning of an interaction. Mr Gough showed how the interpretation of a significant interaction is facilitated by plotting the experimental results against the main effects, inserting limits as on a control chart, and studying the disposition of the experimental points relative to these limits. Further, if similar experiments are repeated many times, their results and the results of their analysis can be plotted in control chart form to demonstrate their stability and hence their value as a basis for action.

GLASGOW GROUP

4th March 1953. Probability and Statistics: B. Woolf

9th April 1953. Statistics in the Transport Industry: T. Irvine Kinlay

LEICESTER GROUP

12th March 1953. The Uses of Snap Observation Methods in a Factory: K. Boyd

23rd April 1953. Annual General Meeting

LONDON GROUP

6th March 1953. Certain New Sequential Procedures: S. Rushton

After briefly outlining the main features and consequences of sequential procedures Dr Rushton gave an account of such procedures for certain tests where composite hypotheses are involved: the one-sided and two-sided sequential *t*-tests, tests for the equality of variances of two populations, and sequential procedures for the analysis of variance. In each case the procedure was based on the appropriate likelihood ratio with the usual Wald-Barnard sequential limits $\beta/(1-\alpha)$ and $(1-\beta)/\alpha$ for specified risks of error α and β. Dr Rushton referred to two previous papers (*Biometrika*, **37**, 1950, p. 326, and **39**, 1952, p. 302) in which the detailed application of the sequential

t-tests had been discussed. Simplified approximate procedures for these tests, not involving the use of tables of exact critical limits, proved quite satisfactory in practice. Suitable limits for the expected sample size could be obtained, and the actual distribution of the sample size and the excess of the likelihood ratio over the boundaries when a decision is reached, were illustrated by diagrams based on sampling experiments. The sequential procedures suitable for testing hypotheses in the analysis of variance depend on the particular models assumed. Dr Rushton described the scheme for an analysis of variance based on a one-way classification. Procedures for testing whether two populations have equal variances can be based either on the usual sample estimates of the variances or on the ranges of subsamples of a given size. If the sample estimates of variance are used and the population means are unknown, the resulting loss of information is, as usual, equivalent to the reduction of the sample sizes by one item. Tables of exact critical limits for these tests and also tables to facilitate analysis of variance tests are to appear with explanatory material in *Sankhyā*.

10th April 1953. A Statistical Department in the Gas Industry: J. Reece

The growth and the use made by the Board's management of the Market Survey Section of the North Thames Gas Board was described. Its work was illustrated by description of three typical pieces of research, namely (i) an investigation into the factors determining the demand for gas, (ii) the effect of climatic temperature on gas consumption, and (iii) the accuracy of gas meters.

By attaching the section directly to the management it became a centre of operational research linking the problems of the commercial, engineering, technical research, and other departments. For example, the study of the effect of air temperature on gas consumption was of assistance in planning repairs and maintenance for times of low demand, in forecasting demand during cold spells, in partitioning known consumption over a period between parts of that period, and for co-ordinating sales policy to production.

1st May 1953. The Utility and Futility of Statistics illustrated by Papermaking Research: A. F. Tout

Mr Tout's theme was that, especially in the early stages of a study, the calculation of correlation coefficients, regression lines, and analyses of variance, are often unnecessary, since the existence of relations and significant effects can usually be judged from the results themselves, especially if graphical methods are employed to display them. He found some use both for partial correlation, which was usually only revealed by calculation, and for calculation of a correlation where a small effect is being sought, as, for example, when the product is only just good enough for its purpose and small changes of quality are important. Examples were given of statistical reasoning applied to papermaking studies and how the principles of experimental design were needed even when the analysis was not carried out in full.

MERSEYSIDE GROUP

24th March 1953. Present Day Practice in the Control of Quality in the USA: E. D. van Rest

21st April 1953. The Evaluation of Detergent Performance in Domestic Dishwashing: C. Pugh

See the article in this issue of *Applied Statistics*.

NORTH-EASTERN GROUP

18th March 1953. Self-protection in Purchasing; Practical Sampling Schemes: W. J. Ross

15th April 1953. Some Statistics of Interviewing: N. D. Squirrel

SHEFFIELD GROUP

26th March 1953. The Choice between Alternative Statistical Tests: E. S. Pearson

The considerations which should govern the choice of a statistical test were brought out by detailed discussion of two examples. Professor Pearson's first example was drawn from the familiar field of the analysis of variance for a one-way classification. Here the solution was clear-cut, largely because the most likely alternatives to the null hypothesis could be readily specified.

The second example concerned the problem of testing whether a sequence of events following one another in time or space could be regarded as occurring at random. Familiar tests of randomness could be derived either from the exponential distribution of intervals between events or the Poisson distribution of counts of events occurring in fixed periods, but these tests were not wholly satisfactory because they did not take full account of the order of the intervals or counts. Another interesting method of approach was to regard the events as an ordered set of observations drawn from a rectangular distribution and to consider their displacements from the 'expected' positions. This led to use of the ω^2-test of Smirnoff and to the test of the significance of the maximum displacement, due to Kolmogoroff. However, the problem of the most efficient test could hardly be solved unless it was possible to specify the form of departure from randomness that it was appropriate to look for in each particular situation.

30th April 1953. Some Experiments involving Subjective Judgment: W. N. Jessop

This paper discussed experiments in which articles are presented to a panel of judges and preferences recorded. Two cases arise, one in which each judge ranks the complete set of articles, and the other in which the articles are presented to the judge in pairs and independent judgments made. Both cases are illustrated with examples from the textile field and particular attention is given to the method of paired comparisons, alternative ways of analysing the results being compared and criticised. A comparison is made of the results obtained from two groups of judges with the same articles and results from the same judges on separate occasions. The coefficients of consistency of individual judges on the two occasions are compared.

SOUTH WALES GROUP

19th March 1953. Some Administrative Problems of a Statistical Department: R. F. George

23rd April 1953. The Application of Statistical Methods to Problems of Mining: B. H. P. Rivett

STUDY SECTION OF THE ROYAL STATISTICAL SOCIETY

Reports of Meetings

LONDON

11th February 1953. Money Saving and Money Making by Statistical Methods: W. Buckland, J. Murdoch, T. Cauter, and G. Bennett

This meeting was designed to foster the 'public relations' of statistics. The four speakers each gave a case history of a project which had resulted in the saving of a large sum of money. The examples covered transport, factory design and materials handling, marketing, and telephone services. Dr Buckland outlined an experiment using a 'cross-over' designed experiment concerning the viscosity of oils in motor transport relative to the length of the journey and other factors. Mr Murdoch described how statistical methods led to the reorganisation of an internal transport system and how such reorganisation avoided the necessity for the laying down of a new additional system and the consequent redesign of the factory and loss of floor space. Mr Cauter told about a study of marketing using a balanced sample and linking this with advertising in a restricted territory. Mr Bennett explained the use of sample surveys which enabled the GPO to save many thousands of pounds on the issue of telephone directories. By the use of a small sample it was possible to predict the likely increased demands upon the directory inquiry service should the directory be permitted to get out of date.

11th March 1953. The Possibilities and Limitations of Electronic Computors in Statistical Calculations: J. A. C. Brown

Mr Brown explained that electronic calculators were most useful in the solution of problems which were amenable to an iterative method. Where the amount of data to be fed into the machine was large but the amount of calculation necessary was small, it was generally not economic to consider their use. The best procedures were those which took into account the special characteristics of various types of mechanical aids which were available. On a recent problem he had found it to be economic to carry calculations as far as gross sums of squares using the more usual punched card equipment, but to transfer to the electronic calculator the solution of the equations which were required. Electronic calculators were particularly suited to the solution of linear equations. It was pointed out in the discussion that the cost of obtaining solutions of n linear equations from the National Physical Laboratory's machine was $£n^2/50$, with a minimum of $£2$. For matrices of reasonable size it not only became quicker to make use of electronic machines, but also cheaper. Discussion was also concerned with the size of matrices which might be solved, with a series of contributors each in turn quoting a larger one until the chairman suggested an end to the 'bidding' when the score was 134×134!

8th April 1953. A Survey of Forecasting of Economic Indices: E. Shankleman

A summary of this paper appeared in the June 1953 number of *Applied Statistics*.

19th May 1953. Sample Surveys and Administrative Action: L. Moss

The speaker, who is the Director of the Government Social Survey, outlined a large number of projects which had been carried out by his unit within recent years. His main thesis was that there was little or no difference between social research of an 'operational' character and 'pure' research. The former approach could lead to findings of importance to the latter, and the latter could lead to social action. This unity was, he thought, quite often not realised, and this was to the disadvantage of both types of research.

The basic material of this paper has since appeared in the *International Social Science Bulletin*, 1953, **5**, 2, and copies may be had on application to the Social Survey.

BOOK REVIEWS

C. Radhakrishna Rao: *Advanced Statistical Methods in Biometric Research.* New York: John Wiley and Sons; London: Chapman and Hall, 1951. Pp. 390. 60s.

One feature of the rapidly expanding science of statistics is the large number of text-books which have been published, particularly in recent years. Most of these are written at an elementary level, often with a particular field of application in mind, and a high proportion are orthodox and similar to each other in essential matter.

Professor Rao's book is not 'just another' of this type, and represents a valuable addition to the range. It is essentially a book on mathematical statistics and requires a knowledge of mathematics at least up to university degree standard for its understanding. The book is therefore intended for the serious student of statistics and for the professional statistician. The development of the subject is quite general and not limited to the field of biometric research except in the examples used to illustrate the methods.

The mathematical treatment is severely algebraic, and the use of vector and matrix algebra in particular have considerably simplified the presentation; this results in elegance and a saving of space, and in a surprising amount of information being presented in a book of no more than normal size. It is unquestionably a good book and ranks with the other leading books on mathematical statistics.

The first chapter gives a condensed version of the algebraic methods used in the book and serves as a very useful 'refresher' to the reader. The second chapter is also in the nature of an introduction, and combines the uses of the mathematical and statistical tools. This chapter deals mainly with various distributions, distribution of linear functions, quadratic forms, multivariate distribution, and least squares. The treatment, although brief, is adequate and up to date. Some of the topics treated comprehensively in the remainder of the book are linear estimation, tests of linear hypotheses, statistical inference, the general theory of estimation, homogeneity of variances and correlations, multivariate analysis, and discrimination between populations. The biometric examples used to illustrate the methods are well chosen, but some of the numerical methods used, e.g. for inverting matrices, although reasonably satisfactory, are not now considered to be the best, particularly when dealing with a large number of constants.

Statistical methods is still a rapidly expanding subject, and books dealing with the advanced aspects tend to become obsolete. It is hoped that Professor Rao will, by means of further editions, keep the subject matter up to date. One such instance of this need already appears in the section dealing with the tests on the homogeneity of variances: Dr G. E. P. Box has shown, in a paper which will appear shortly in *Biometrika*, that the current tests on the homogeneity of variances are unsatisfactory and frequently misleading. No doubt Professor Rao will attend to this matter in the second edition of his book.

O. L. DAVIES

A. Hald: *Statistical Theory with Engineering Examples.* Pp. xii + 783. 72s. *Statistical Tables and Formulas.* Pp. 97. 20s. New York: John Wiley and Sons; London: Chapman and Hall.

The engineer uses statistical technique exactly in the same way as a machine tool. While dictating a mode of thought such use is a process leading to a product whether applied to production or research, and any book on statistical methods must be judged by the engineer on the assistance it will give him in the understanding and use of that process and by the readiness of that assistance.

Professor Hald's *Statistical Theory* must be regarded as an information store rather than an immediately available tool. The background theory is very comprehensive, but in general it is not readily applicable by the technician. This is not a book for the beginner, and the average practising engineer could not refer to a particular section at short notice. On the other hand, given a mathematical training and experience in statistical methods, the reader will find an extremely good coverage in this book which would make a valuable addition to his library. It is a book for the statistician dealing with engineering problems rather than for an engineer dealing with statistical investigations.

Certain chapters are very good indeed. Those on the normal distribution and fractile diagram, the *t*-distribution, the analysis of variance, and the design of experiments deserve special mention. An outstanding feature of the book is the extensive bibliography, copious references being given in each section not only to parallel literature but also to technical examples. Some treatments are heavy. Thus producer-consumer risks could be discussed with greater clarity and speed by graphical illustration of two distributions. The enormous number of symbols used, while entirely logical and scientifically designed, demand a good memory.

The book is based on Professor Hald's Danish textbook and in the dozen or so numerical examples from engineering practice only metric units are used, whereby much of their value for British engineers is lost. To take one case, the English-speaking engineer thinks of cement strengths in pounds per square inch, and a discussion in the unfamiliar units of kilogrammes per square centimetre makes just another academic example.

In a book of nearly 800 pages the absence of misprints is noteworthy. In fact only one was found. The diagrams are clearly the work of a draughtsman and the printing is excellent.

The companion *Statistical Tables* summarise the important formulae derived in *Statistical Theory* and present complete tables for use with these formulae. References to more extended tables are also included.

F. C. TURNER

Sir Ronald A. Fisher and Frank Yates: *Statistical Tables for Biological, Agricultural and Medical Research.* Edinburgh and London: Oliver and Boyd, 1953. Pp. xi + 126. 21s.

The new edition of these well-known and important tables differs from the third by the addition of five new tables and a revision of a small part of the introduction. That this collection of tables is well known is shown by the fact that it has gone into a fourth edition. That it is important arises from the facts that it contains some tables not otherwise readily available, and that it is comprehensive enough to serve most of the needs of a wide range of users of statistical methods—a range extending far beyond that suggested by the title. Only about one-tenth of the tables have an application restricted to biology, agriculture, and medicine. L. H. C. TIPPETT

J. Mounsey: *An Introduction to Statistical Calculations.* London: English Universities Press, 1952. Pp. 351. 15s.

Students and teachers of economic statistics, particularly the latter, will appreciate the copious supply of exercises and examples given in this book. Some are original, some are drawn from intermediate examination papers, and some are taken from the Royal Statistical Society's own Certificate papers. The examples are linked by a text which is more than a list of formulae and less than an exposition. While there is always a risk that some students, misled by the simplicity of presentation, may use the book as an inadequate substitute for an exposition, there is no denying the usefulness of some of the reminders and warnings to other students using it as a legitimate help in revision.

As often happens in introductory books, the book is at its best in early sections relating to frequency distributions, index numbers, and moving averages. Here the exercises involve interesting data, often drawn from official sources. The book's quality deteriorates when it comes to deal with more advanced topics such as probability and significance tests. There is a serious error on p. 170, where the expression $\sigma_y\sqrt{(1 - r^2)}$ is described as 'the standard error of a regression line.'

On the whole the exercises are such as to test technique, not understanding, and in the later sections of the book tend to lack the atmosphere of reality. Even so, Mr Mounsey will save many a teacher many a headache from hunting for material to illustrate lectures or lessons. G. H. JOWETT

MADAS

Fully Automatic Calculator

Model 20ATG, illustrated above, is the ideal machine for every kind of statistical work.

It has four visible registers: two with 20-figure dials showing single and accumulated products and two 10-figure dials for multipliers or quotients. As individual products appear in the lower product dials they are automatically accumulated in the upper product register; similarly multipliers and quotients can be accumulated in the right-hand 10-figure dials, while the left-hand register automatically clears after each calculation.

It is possible to calculate in all registers positively or negatively or in any combination required.

Among a number of exclusive features it embodies the transfer device and squaring feature without resetting.

These features, together with the accumulating registers, obviate the necessity of writing down intermediate answers.

Write for full details or free demonstration to

THE MULDIVO CALCULATING MACHINE CO. LTD.

49 Queen Victoria Street, LONDON, E.C.4

Telephone: CITY 3447–8

and at

Birmingham, Manchester, Leeds, Newcastle, Glasgow, Dundee, Belfast, Dublin

Journal of the Royal Statistical Society

SERIES A—GENERAL

Price Each Part 15s.

Subscription Post Free £3 1s.

Principal Contents of Volume 116 Part III 1953

Construction of an Index Number of Gilt-Edged Values (with Discussion) .. .. *M. S. Rix*

Statistical Investigations of Labour Productivity in Cotton Spinning (with Discussion) *L. H. C. Tippett* and *P. D. Vincent*

The Treatment of Errors .. .. .. *Dudley Seers*

The Industrial Wages Bill and Earnings of Operatives in 1935 and 1948 .. .. *H. Leak*

Criminal Statistics .. .. *T. S. Lodge*

Timber Statistics .. .. *J. J. MacGregor*

Reviews of Books, Statistical and Current Notes, Periodical Returns

ROYAL STATISTICAL SOCIETY

4 Portugal Street

London - W.C.2

Journal of the Royal Statistical Society

SERIES B—METHODOLOGICAL

Price Each Part
22s. 6d.

Annual Subscription
Post Free 45s. 6d.

Contents of Volume XV No. 1 1953

Sequential Estimation (with Discussion) *F. J. Anscombe*

Statistical Inference (with Discussion) *D. V. Lindley*

The Random Division of an Interval—Part III *P. A. P. Moran*

Simple Graphical Techniques for Calculating Serial and Spatial Correlations, and Mean Semi-squared Differences *G. H. Jowett* and *J. F. Scott*

On the 'Transition Probabilities' corresponding to any Accident Distribution *J. O. Irwin*

A Note on the Time-Homogeneous Birth Process *P. Armitage*

Application of the Theory of Stochastic Processes to the Study of Irreproducible Chemical Reactions and Nucleation Processes *K. Singer*

Goodness of Fit Tests for Simultaneous Autoregressive Series *M. S. Bartlett* and *D. V. Rajalakshman*

The Analysis of Multiple Stationary Time Series *P. Whittle*

An Approximate Treatment of the Properties of the Correlogram and Periodogram *J. D. Sargan*

ROYAL STATISTICAL SOCIETY
4 Portugal Street
London - W.C.2

THE

MARKET RESEARCH SOCIETY

Limited (by Guarantee)

President: SIR ARTHUR BOWLEY, C.B.E., Sc.D., D.Litt.

The Market Research Society is a professional organisation for all those using field survey techniques for market, social, or economic research. A regular programme is organised of monthly luncheon meetings, discussion groups on technical subjects, and week-end courses in market research. The Society offers a prize in the Final Examination of the Association of Incorporated Statisticians.

Membership is open to all engaged or interested in this field of research, whether in business, administrative, or academic spheres.

The Society was incorporated under the Companies Act, 1948, on 16th April, 1953.

Registered Office:

73 CHEAPSIDE, LONDON, E.C.2

Second edition, revised and greatly enlarged

SAMPLING METHODS FOR CENSUSES AND SURVEYS

F. YATES, F.R.S.

This well-known book, written for those with little previous training in mathematical statistics, covers all the modern developments of sampling theory which are of importance in census work—whether sociological, agricultural, or economic. It gives also an adequate discussion of problems met with in the planning, execution, and analysis of sample surveys, the treatment in general being much fuller than is found in textbooks on statistics or in courses on the subject. This second, revised edition amplifies certain aspects in the light of advances in the last four years and devotes a chapter to the complexities arising in the analysis of investigational surveys, as distinct from surveys of the census type, with instructive examples from recent work.

Pp. xiv + 420 (9in. × 6in.). 38s. net

CHARLES GRIFFIN & CO. LTD
42 DRURY LANE, LONDON, W.C.2

U.S. publication rights held by
The Hafner Publishing Co.
31 East 10th Street
New York, 3

Agents for India, Ceylon and Burma:
The Progressive Corporation
51 Esplanade Road, Fort
Bombay, India

PITMAN

Statistics for Professional Students

By R. L. A. Holmes, B.Com. Provides an excellent introduction to the study of statistics, and is specially designed to appeal to those students preparing for examinations in which a knowledge of this subject is required. In this new Second Edition most of the examples have been redrafted to bring them into line with post-war statistics. 15s. net.

Modern Elementary Statistics

By John E. Freund. Forms a general introductory course in statistics whose primary aim is to familiarise students with the ideas and concepts which are fundamental to an understanding of statistical methods. 40s. net.

Parker St., Kingsway, London, W.C.2

BINDING

We have pleasure in announcing that we are in a position to carry out various styles of binding according to the wishes and requirements of individual customers.

Cloth binding cases are now available for *Applied Statistics*, Volume II.

The rebinding of damaged volumes is undertaken.

Facilities are also available for the special binding of Presentation and de Luxe editions, Visitors' Books, Rolls of Honour, Illuminated Addresses, Antiquarian books, etc.

Quotations for all types of binding work will be gladly given on application to

OLIVER AND BOYD LTD.

BOOKBINDERS

14 HIGH STREET, EDINBURGH, 1

APPLIED STATISTICS

INDEX

accountant and statistician, 138
'Advanced Statistical Methods in Biometric Research,' review, 202
agricultural returns, sampling, 152
analysis of variance, quality control and, 197
auditing, use of samples, 180

baby, first, 140
Bartlett, M. S., 44
Beech, D. G., 73
Beer, S., 160
Benjamin, B., 140
Bennett, G., 200
Billewicz, W. Z., 166
'Biological, Agricultural, and Medical Research, Statistical Tables for,' review, 203
'Biometric Research, Advanced Statistical Methods in,' review, 202
Boddington, A. L., 70
book reviews, 70, 71, 72, 202, 203, 204
Booker, H. S., 71
Bray, F. S., 138
British Standards Institution, 67
Brown, J. A. C., 200
Buckland, W., 200
business barometers, 86
business forecasting, 101
Butler, J. D., 24

Cauter, T., 200
census of distribution, 1
census, one per cent. sample, 138
Centre de Formation des Ingénieurs et Cadres aux Applications Industrielles de la Statistique, 196
Champernowne, D. G., 118
chemical manufacture, problem in, 30
Clark, D. A., 1
Committee on Statistics in Industry and Technology (ISI), 135
communication engineering, stochastic processes in, 49
control charts, massed, 160
Corlett, T., 193
correlation analysis:
  experiences, 73
  use for estimating factor of proportionality, 79
  use for estimating importance of factors, 83
  use for testing theoretical relationship, 75
correlation of time series, 137
cotton spinning mill, experiment in, 184

Davies, O. L., 70, 136, 202
defectives, sequential sampling, 118
'Design and Analysis of Experiments,' review, 70
detergent performance, 172
Devons, E., 72
distribution, census of, 1
domestic dishwashing, 172
Downton, F., 30

economic forecasting, 86
electronic computors, 200
'Engineering Examples, Statistical Theory with,' review, 203
epidemics, stochastic processes and, 56
expenditure of individuals, estimation from family totals, 131
exports, volume of net, 166
extrapolation, dangers, 74

factor analysis, origin and development, 107
Fisher, R. A., 203
forecasting:
  business, 101
  economic, 86
Frances Wood Memorial Prize, 134

Garvie, A., 140
gas industry, 198
glass manufacture, 136
Gough, V. E., 197
Grebenick, E., 138

Hartley, H. O., 20
Hughes, C. D., 101

Ilersic, A. R., 70
index of net exports, 166
individual expenditure, estimation from family totals, 131
Industrial Applications Section of the Royal Statistical Society, 68, 136, 197
industry, selling statistics to, 136
influenza, spread within households, 140
inspection, 145
International Statistical Institute, 135
'Introduction to Statistical Calculations,' review, 204

Jessop, W. N., 199
'Journal of Industrial Economics,' 67
Jowett, G. H., 137, 204
jurors, proportion of, 193

Keen, J., 13
Kempthorne, O., 70
Kemsley, W. F. F., 132
Kendall, M. G., 71, 197

Lamberth, D. L., 139
Lloyd, F. J., 140

McKeon, A. J., 39
Martin, W. J., 66
Mason J., 138
measles epidemic, stochastic model, 56
medical records, 140
medical statistics, nature and sources, 140
medicine, stochastic processes in, 56
'Methods of Statistics,' review, 72
money benefits of statistical methods, 200
Monte Carlo method, 47
morbidity data, 66
Moroney, M. J., 136
mortality in Scotland, 140
Moss, L., 201
Mounsey, J., 204
Muir, A., 30
Murden, K. G., 140
Murdoch, J., 200

national income forecasts, 90
Nicholson, J. L., 132
Nuttall, C., 138

Page, D. J., 13
Palca, H., 152
paper-making research, 198
Peake, R. E., 184
Pearson, E. S., 199
petroleum consumption, short-term fluctuations, 24
planning an experiment, 184
post and mail, 138
prediction, statistics and, 101
Price, O. O., 39
proportionality, estimating factor of, 79
Pugh, C., 172

quality control:
  and the analysis of variance, 197
  in the USA, 141
queueing problems, 50

ranking, 197
Rao, C. Radhakrishna, 202
Reece, J., 198
'Revue de Statistique Appliquée,' 196
Rushton, S., 197
Russell, M., 140

samples:
  use in auditing, 180
  use of proportion of jurors in design, 193
sampling:
  of agricultural returns, 152
  economics of sequential, 118
sampling (*continued*):
  serial number digit compared with systematic and random, 39
schedules, design, 139
sequential procedures, 197
sequential sampling, economics, 118
serial number digit sampling, 39
Shankleman, E., 86
sickness absence statistics, 140
Social Survey, 201
'Sources and Nature of the Statistics of the United Kingdom,' review, 71
'Statistical Calculations, an Introduction to,' review, 204
'Statistical Tables' (Hald), review, 203
'Statistical Tables for Biological, Agricultural, and Medical Research,' 203
statistical tests, choice, 199
'Statistical Theory with Engineering Examples,' review, 203
'Statistics and their Application to Commerce,' review, 70
statistics and prediction, 101
Stein, L., 140
stochastic processes, 44
Study Circle on Medical Statistics, 140
Study Section (of the RSS) forecasts, 95
Study Section of the Royal Statistical Society, 138, 200
subjective judgment, 199
successive differences, use for estimating variability, 13
Sutherland, I., 140

telephone lines, delays, 65
theoretical relationship, testing by correlation analysis, 75
Thomson, A., 140
time series:
  correlation of, 137
  stochastic processes, 52
time variations, variability estimated from successive differences, 13
Tippett, L. H. C., 72, 203
Tout, A. F., 198
trends, elimination, 136
Truesdale, S. P. L., 136
Turner, F. C., 203

van Rest, E. D., 66, 72, 141
variability, estimation from successive differences, 13
Vincent, D. F., 107

Williams, N., 180
Wood, S., 138

Yates, F., 203
yield in chemical manufacture, 30

Printed in Great Britain at THE KYNOCH PRESS, *Birmingham*